N465

D-S
2123

DEPOSITORY

22 AUG 1975

NAVAL RESEARCH LOGISTICS QUARTERLY

JUNE 1975
VOL. 22, NO. 2

OFFICE OF NAVAL RESEARCH

NAVSO P-1278

407-B

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Murray A. Geisler
The RAND Corporation

W. H. Marlow
The George Washington University

Bruce J. McDonald
Office of Naval Research

MANAGING EDITOR

Seymour M. Selig
Office of Naval Research
Arlington, Virginia 22217

ASSOCIATE EDITORS

Marvin Denicoff
Office of Naval Research

Alan J. Hoffman
IBM Corporation

Neal D. Glassman
Office of Naval Research

Jack Laderman
Bronx, New York

Thomas L. Saaty
University of Pennsylvania

Henry Solomon
The George Washington University

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Subscription Price: $10.00 a year in the U.S. and Canada, $12.50 elsewhere. Cost of individual issues may be obtained from the Superintendent of Documents.

The views and opinions expressed in this Journal are those of the authors and not necessarily those of the Office of Naval Research.

Issuance of this periodical approved in accordance with Department of the Navy Publications and Printing Regulations, P-35 (Revised 1-74).

DYNAMIC MULTIATTRIBUTE MODELS FOR MIXED MANPOWER SYSTEMS

A. Charnes

University of Texas

W. W. Cooper

Carnegie-Mellon University

R. J. Niehaus

Department of the Navy, Office of Civilian Manpower Management

ABSTRACT

This paper reviews a wide variety of manpower and personnel models of the goal programming variety. This is done from a strategy-oriented point of view addressing the problems of interest for immediate implementation as well as basic problems of manpower model research development. Particular emphasis in this paper is concerned with how analytical models can be brought to bear on the problems of combining military and civilian manpower into one management system. This includes a discussion of the computer support arrangements necessary to implement the models. First, we discuss an extension of multilevel models to provide an integrated approach to program planning which includes the dynamics of the manpower requirements-inventory relationships of mixed military-civilian manpower systems. Then, focus is given to some of the potential Navy applications particularly in terms of ways the outputs from the global multilevel model might be interfaced with assignment models for operational planning. The paper concludes with a discussion of static and dynamic multiattribute assignment models which operate on the individual man-job matching level. It is at this level of detail that dynamic mixed manpower systems might be constructed for use in equal employment opportunity planning and for local organization design studies.

INTRODUCTION

A mixed manpower system contains two relatively distinct classes of manpower which must be integrated for overall management purposes. This is the case with the U.S. Navy which contains over 850,000 personnel roughly divided into two-thirds military and one-third civilian personnel.* In this paper we will be discussing how analytical models can be brought to bear on the problems of combining military and civilian manpower into one management system. This will be done from a strategy-oriented point of view addressing the problems of interest for immediate implementation as well as basic problems of manpower model research development.

One of the backdrops for this paper will be a review of the research underway since 1967 by the Management Sciences Branch, Office of Civilian Manpower Management (OCMM) jointly with the

*See [2] for a discussion by VADM D. H. Bagley of the future of the Navy personnel structure.

Center for Cybernetic Studies, the University of Texas and by the Management Sciences Research Group, Carnegie-Mellon University. Another backdrop will consist of a review of some existing bodies of knowledge that might be useful in this research. In this paper we will be looking toward the problems associated with (1) civilian manpower systems, (2) military manpower systems, and (3) mixed manpower systems. Finally, then, we shall delineate some recent additional developments in areas like manpower assignment and modeling for officer distribution planning.

BACKGROUND AND RESEARCH STRATEGY

The OCMM manpower research has its roots in basic research supported by the Office of Naval Research. This work is also oriented toward the practical problems in which it is grounded, and hence it is both basic and applied. The current and planned OCMM research and development program was initiated as a result of a policy decision of the Assistant Secretary of the Navy (Research and Development). It is part of the U.S. Navy applied research programs leading to Advanced Development and ultimate implementation. Support for the work is provided by the Chief of Naval Development in coordination with the Personnel and Training and the Operations Research Programs of the Office of Naval Research. The Advanced Development phase is being supported under the guidance of the Navy Personnel Research and Development Center.

Being both basic and applied in its intent, the OCMM research program has a variety of goals. Among these goals, the OCMM research program seeks to extend the dialogue concerning models for studying military-civilian interactions. This was among the objectives of the Assistant Secretary of the Navy (Manpower and Reserve Affairs) in his endorsement of this research. This paper should be viewed as an initial analytical statement of some of the possible models which could be used in such integrated military-civilian planning. But this is not the only such effort. Research in this area is also being undertaken by analysts in the Office of the Deputy Chief of Naval Operations (Manpower and Naval Reserve). This includes their studies in coordination with the Operations Research Center of the University of California, Berkeley which is supported by the Office of Naval Research. Their emphasis to date has been with military manpower systems.*

Much of our past work has been with aggregative civilian related models. It has covered all levels of manpower decision making. This past work can be discussed in terms of a series of phases as follows: First, it was necessary to identify and formulate the problems in a way that would produce operational results. This could not be accomplished within the state of the existing modeling art. Hence, in the initial model development it was found necessary to extend the literature of Markov decision processes and combine these developments with goal programming.† This was done in order to obtain models that could be suitably addressed to the aggregative recruiting requirements problem in the context of the U.S. Navy's civilian work force. In brief, these models consist of embedding a series of Markov processes in a goal programming context. This then made it possible to deal with the dynamics of manpower recruitment and planning in ways that tied directly to the context of prescribed manpower "end strengths" or requirements. These end strengths are the prevalent guide for these planning decisions in Navy practice. The term "end strength" which was interpreted as a guide rather than as

* See [4.1] for a review of manpower modeling issues and trends and [12] for a survey of currently operational Navy manpower planning systems. In addition, the Navy Personnel Research and Development Center and the Air Force Military Personnel Center have surveys underway of existing manpower modeling efforts.

† See chapter X of Charnes and Cooper [3].

a strict upper limit, could be accommodated by the goal programming aspects of the model. Transitions of recruits and job incumbents from one position to another could be dealt with by stochastic elements of a Markov process. This could be done after the Markov processes were extended to allow for alterations from period to period due to such activities as recruitment, etc., in prior periods. To handle all of this the model had to be extended to so-called "decision Markov processes" which were embedded in a goal-programming context, thereby extending both areas of modeling, as already indicated.

The above description refers only to the manpower aspects of these developments. What was wanted, however, was a model that addressed financial budgeting and manpower planning (i.e., dollars and men) as twin aspects of a simultaneous decision process. After this was done the models were then extended to include training and relocation possibilities. This included probabilistic controls (or risk variations) of a chance-constrained programming variety and with varying probabilities, of so-called quasi-Markov form so that, again, this modeling was brought to a new state of development.

Finally, these recruiting requirements models were extended to make contact with other Navy modeling efforts. This was done by means of a series of multilevel models structured to integrate manpower planning with related resource requirements planning. A multilevel model might be defined as one in which more than one level of decision making is included in the same model. This was done so that the results could be used in overall program planning where such resource requirements had to be considered along with associated manpower and financial aspects of such plans. These program planning extensions were effected in liaison with the Resource Analysis Division of the Center for Naval Analyses (see [4.1] and [4.5]).* The multilevel models were also extended to include risk variations by means of the use of chance constrained programming [4.5] so that in this way the models could be structured to include probabilistic variations explicitly in pricing out variable force mixes with associated manpower variables.

The aggregative civilian recruiting requirements models are well into implementation. This includes an extensive computer driven system to support large-scale prototypes. This computer software has been named the Computer-Assisted Manpower Analyses System (CAMAS).† Most of the prototypes to date have dealt principally with the manpower dynamics aspects of the problem. However, small-scale prototypes have been completed that demonstrate the computational feasibility of the multilevel models. This includes a prototype that uses manpower transition data and resource related input-output coefficients from existing Navy information systems. The CAMAS software is designed to provide data on historical trends, input to most large linear programming packages, and output a wide variety of management reports. We have also developed a conversational interactive computerized version of the models in association with G. L. Thompson [11] of Carnegie-Mellon University.

Following the recommendations of the Director, Research, Development, Test, and Evaluation of the Chief of Naval Operations and at the request of the Naval Analysis Branch of the Office of Naval Research and of the Officer Distribution Division of the Bureau of Naval Personnel (BUPERS), work has been started on officer planning models. This research for the present is concentrated on the personnel management aspects (i.e., detailing, assignment, etc.) rather than on the manpower planning aspects of the programming phase of the PPB cycle such as is included in the multilevel models. Finally, with

* See J. Augusta [1] and N. Hibbs [9] for a description of the Navy Requirements Model (NARM) that the Center for Naval Analyses has developed for the General Planning and Programming Division of the Chief of Naval Operations.

† See R. J. Niehaus [4.1], L. Mannis [10], and M. Gerwitz, S. Korn, and R. J. Niehaus [6] for descriptions of the types of analytical studies possible with CAMAS.

the assistance of BUPERS, goal-programming models are also being developed for multiperiod officer rotation and career profile analyses.†

As the above discussion has already indicated, this applied work required a series of accompanying developments in altering the states of modeling and personnel practice. This includes suggested changes in other practices such as program planning and financial management. Fundamental research has also been undertaken in areas such as dynamic multiattribute assignment models [4.6]. This work was done in anticipation of subsequent extensions to organization design, especially dynamic organization design. Mathematically, this has included a spectral analysis approach to obtain mixtures of distributions for career paths in relation to employee aspirations and organizational goals, and a generalization of the eigenvalue problem to deal with recruitments and personnel assignments used to vary the Markov elements themselves.

A broad-gauged applied research program has now been put together as a result of the policy decisions of the Assistant Secretary of the Navy (Research and Development) and the Assistant Secretary of the Navy (Manpower and Reserve Affairs). Emphasis in this research is in the development of the dynamic multiattribute assignment models. However, these personnel management oriented models are also to be synchronized with overall planning or global models. Without such synchronization little impact can be made on the budget preparation phase or structuring of the tasks to which the personnel must be ultimately assigned. Our research is aimed toward adapting or altering the preceding original dynamic multiattribute extensions of Charnes, Cooper, Niehaus, and Stedry [4.6] in order to bring this work into contact with the extant literatures and related bodies of fact, experience and theory in sociology and psychology. This is being done with B. Moore of the University of Texas.**

In conjunction with the above research program we have continued our research into algorithms as well as model development. In particular, working with D. Klingman of the University of Texas, we have now developed a quadratic programming algorithm which replaces a previously developed approximate one with an exact and constructively efficient linear programming equivalent to the ordinary assignment model. This will allow the eventual computation of very large (i.e., one containing 500,000 variables) assignment problems. See [5] and [8].

As can be observed from the above, our research program encompasses the spectrum of applications, model development, and computer-assisted techniques concerned with manpower and personnel planning. This includes the management practices along with the career management and financial management aspects which can be expected to interact with such applied research. We are continuing our research into global aggregative models along with our investigations of dynamic multiattribute assignment models. This is preparatory to taking account of man-job interactions as elements for dynamic organization design models. In a related, but different direction the work of G. L. Thompson involving conversational interactive computer approaches to these problems is also being continued. It appears that this research has direct applicability to military as well as civilian manpower systems. The remainder of this paper will be concerned with conceptual schemes for integrating global or overall planning models with dynamic assignment or individual man-job models. This will be done in the context of mixed manpower systems.

---

† See D. Cass, A. Charnes, W. W. Cooper and R. J. Niehaus "A Program for Navy Officer Distribution Models," R.R. 349, Revised, Graduate School of Indust. Admin., Carnegie-Mellon Univ., Jan. 1975, and "A Multi-Page Goal Programming Algorithm for Navy Officer Rotations," R. R. 350, Revised, Graduate School of Indust. Admin., Carnegie-Mellon Univ., Jan. 1975.

**See B. Moore and D. Sholtz "Organizational Tests of a Static Multi-Attribute Assignment Model," OCCM Tech. Rpt. 20, and "Using Task Surveys in Assigning People," Civilian Manpower Management No. 4, Winter 1974.

GLOBAL MODELS

We might now begin to relate our work, which we have just discussed, to other areas for potential application. What we will be looking at first will be a system of models for overall or global planning. This will be an extension to our multilevel model to provide an integrated approach to program planning, which includes dynamics of the manpower requirements-inventory relationships of mixed manpower systems. This modeling system is proposed to assist in evaluating the structure of total internal manpower requirements for changing task arrangements, testing policy alternatives as changes occur in the budgetary process, and planning for orderly recruitment or retrenchment over time. As a by-product, data could also be obtained on the projected shortfalls or overages resulting from the budgetary allocation process.

Mixed manpower systems exist in most large organizations. For example, many industrial organizations distinguish between exempt and nonexempt employees. Systemic distinctions are also apparent when planning for the integration of white and blue collar employees. In the case of the U.S. Navy, all the complications of mixed manpower systems are present. Military manpower is assigned centrally while civilian manpower is hired locally. There are a certain number of jobs which can only be filled by military personnel. The stochastic elements of civilian recruitment planning center upon internal mobility and exits. With the advent of the all-volunteer force concept, a troublesome stochastic element in military recruitment planning is guaranteeing an adequate supply of manpower. Thus, the dynamics of the work force impinge differently in the planning for adequate supplies of military manpower.

It appears that in modeling military-civilian planning both the global and local manpower planning problem must be addressed. Military manpower requirements result from the projection of a required operational capability into the future and is a dependent variable in relation to that process. Thus, military manpower requirements are planned centrally with global program requirements playing a fairly direct role. On the other hand, civilian manpower requirements are locally determined one-at-a-time with planning being a reaction to the global requirements. Practical experience indicates that one grand model is doomed to failure. What is needed is an open system of models with room at each decision level for the injection of management judgment.

All of the above is reminiscent of kinds of problems likely to be encountered in industrial manpower planning except for the fact that profit and cost must give way to multiple goal situations in the Navy. The latter, however, is the more general case and conceptual and experimental research has already demonstrated the computational feasibility of constructing a multilevel model for civilian manpower planning.* Thus, we may turn to the latter (i.e., multigoal multilevel models) for discussion with assurance that it will accommodate the former.

Figure 1 shows an extension of the OCMM models to include military as well as civilian planning. In this system, a program planning model with input-output elements and goal programming features is coupled to civilian and military models and hence provides a mixed system. As will be described below, both the program planning and manpower submodels use goal programming. In the case of the manpower submodels, their goal programming structure includes embedded Markov processes just as in the case of the original OCMM models.† Additionally, the objective function of the global model uses a priority weighting scheme to provide the required emphasis among the various program planning

*See Niehaus [4.1] for a comprehensive summary of this research including a discussion of initial prototype experience and computer software designs.

†See [4.1] through [4.5].

		PRODUCER OUTPUT	CIVILIAN ON-BOARD		MILITARY ON-BOARD			
USER/PRODUCER GOAL PROGRAMMING MODEL	PROGRAM PLANNING INPUT-OUTPUT MODEL							FINAL USER GOALS & PRODUCER BUDGETS
CIVILIAN MANPOWER/$ RELATIONSHIPS		cm/$1 cm/$2	−1 −1				= =	0 0
MILITARY MANPOWER/$ RELATIONSHIPS		mm/$1 mm/$2			−1 −1		= =	0 0
CIVILIAN MANPOWER GOAL PROGRAMMING MODEL			CIVILIAN MANPOWER MODEL					CIVILIAN MANPOWER GOALS & ATTRITION CONSTRAINTS
MILITARY MANPOWER GOAL PROGRAMMING MODEL					MILITARY MANPOWER MODEL			MILITARY MANPOWER GOALS & ATTRITION CONSTRAINTS
TOTAL FORCE CONSTRAINTS			IIII IIII		IIII IIII		≤ ≤	TOTAL MANPOWER SPACES

FIGURE 1. Multilevel model for total force planning, 10 Jan '73.

and manpower goals. This done by minimizing the weighted sums of the deviations from the various goals and the relative priorities of hiring (enlisting) for firing (releasing from active duty). This multilevel model obviously is designed to assist in the planning for the requirements of the naval shore establishment.

We will now elaborate on some of these developments in more detail. Figure 2 shows the structure of the input-output model for program planning. In this model the goals are the final user support requirements or amount of services the shore establishment must provide to the Fleet. The α are relative priorities or penalties used to weight the deviations from the goals. The goals are included in the first set of equations which state that the total amount of support furnished each user less the amount over or plus the amount under will be equal to the total requirement or goal. In any solution one will obtain for each equation the level of output associated with the goal and either the overage or underage (or zero deviation if the goal is met right on). This stems from the fact that one cannot be both over and under a goal at the same time. It should be noted at this point that all outputs of the program planning submodel would be expressed in their dollar equivalents so that one can be consistent in adding the outputs for each of the producers.

The next set of equations ensures that the distribution of output from each of the producers (e.g., the major elements of the shore establishments) is in the right proportion to the requirements of the final users. One can simultaneously calculate the support-on-support requirements, or alternatively, as is done here, a second stage allocation process may be used by means of definitional relations. The latter is accomplished by using data from an input-output table which indicate for each producer the proportion of output required to support the final users. This ensures that the final users

	POSITIVE GOAL DISCREP.	NEGATIVE GOAL DISCREP.	FINAL USER SUPPORT	PRODUCER OUTPUT	SIGN	RHS
RELATIVE PRIORITIES	α	α				
FINAL USER GOALS	$-I$ $-I$	I I	I I		$=$ $=$	FINAL USER SUPPORT REQUIREMENTS
SUPPORT USAGE RATES			$-I$ $-I$	I-O RATES I-O RATES	$=$ $=$	0 0
PRODUCTION LIMITS				I I	$\leq$ $\leq$	PRODUCER BUDGETARY LIMITS

FIGURE 2. Program planning submodel that uses input-output analyses.

will be supported at the level required in the model solution and at the same time obtain the total amount of each output which should be produced. These relationships are built into the model by specifying an equation for each final user. This equation states that the sum of the output of each producer multiplied by the proportion consumed by each final user minus the total amount each final user consumes from all producers will equal zero. This forces all the individual pieces of output from each type of support to be in the right proportion to the total.

The next section of the program planning submodel bounds the amount of each kind of output to be produced. An equation is assigned to each producer for each time period which prescribes a limit to the amount that can be produced. This ensures that no producer will exceed its budget in trying to meet the overall goal. At the same time if an excess budget has been allocated to the producer it will show up in the model solution.

Figures 3 and 4 show the civilian and military submodels. In both of these submodels the manpower requirements are set as goals to be obtained. The portions of the objective of the multilevel model which concern these submodels also contains relative priority weights or penalties. These penalty values can be set in a range much lower than those of the program planning submodel so that the program plans have a much heavier weight in determining overall allocations. In the manpower submodels a penalty is paid whenever the manpower requirement for a given category is not met. Relative penalties are also paid for adding to or reducing the work force. The scheme can be set so that reductions have a higher priority than goal deviations or vice versa. Further, these strategies can be mixed so that some of the manpower requirements will be met much more closely than others.

The submodel objectives are subjected to a number of constraints which explain to some extent the differences between the civilian and military manpower systems. In both submodels, the number on board in each job category at the start is set equal to a constant. This ensures that the base period population will be completely accounted for in the solutions. These base period populations are then submitted to a matrix of movement, or transition, rates which distinguishes probabilistically between those staying in a particular job category, those moving to another job category, and those leaving the organization.

At this point the civilian and military submodels begin to differ. In the case of the civilian submodels, provision is made for entries at all levels. In the military submodel, entry is restricted primarily to the lower ranks. This can be done either by allowing entry only at the lower ranks, as indicated on

	POSITIVE GOAL DISCREP	NEGATIVE GOAL DISCREP	ON-BOARD MANPOWER	NEW HIRE MANPOWER	EXCESS MANPOWER	SIGN	RHS
RELATIVE PRIORITIES	β	β		γ	δ		
MANPOWER GOALS	$-I$ $-I$	I I	I I			$\geq$ $=$	MANPOWER REQUIREMENTS
MANPOWER ATTRITION			I $-M\ I$ $-M\ I$	$-I$ $-I$	I I	$=$ $=$ $=$	INITIAL POP 0 0
TOTAL MANPOWER CONSTRAINTS			1111 1111			$\leq$ $\leq$	CIVILIAN MANPOWER AVAILABLE
SALARY BUDGET CONSTRAINTS			(\$/m)1 (\$/m)2			$\leq$ $\leq$	CIVILIAN SALARY BUDGET

FIGURE 3. Civilian manpower submodel.

	POSITIVE GOAL DISCREP	NEGATIVE GOAL DISCREP	ON-BOARD MANPOWER	NEW RECRUITS	RELEASE FROM ACTIVE DUTY	SIGN	RHS
RELATIVE PRIORITIES	λ	λ		μ	ν		
MANPOWER GOALS	-1 -1	1 1	1 1			$=$ $=$	MANPOWER REQUIREMENTS
MANPOWER ATTRITION			1 $-M\ 1$ $-M\ 1$	-1 -1	1 1	$=$ $=$ $=$	INITIAL POP 0 0
MILITARY ONLY REQUIREMENTS			IIII IIII			$=$ $=$	PURELY MILITARY REQUIREMENTS
TOTAL BILLET CONSTRAINTS			IIII IIII			$\leq$ $\leq$	TOTAL BILLETS AVAILABLE
MILITARY BUDGET CONSTRAINTS			$(\$/m)_1$ $(\$/m)_2$			$\leq$ $\leq$	MILITARY PERSONNEL BUDGET

FIGURE 4. Military manpower submodel (Military Model I).

Figure 4, or by setting progressively higher penalty weights on outside entries as one moves up the job categories. Additional constraints are also set for manpower end-strengths and manpower salary budgets. In the military submodel, a constraint also reflects the fact that a certain number of spaces can be filled only by military. This limits the number of spaces which can be traded between military and civilian manpower.

The three submodels are tied together in two ways. First, a set of coupling equations tie the program planning submodel with the manpower models. These are essentially pricing equations which indicate the amount of men per dollar of final user support. Second, an overall manpower space constraint is set across both the civilian and military submodels. This ensures satisfaction of a total manpower space limitation.

The multilevel model differs from a static input-output model in that imbalances and inconsistencies between various goals are allowed and can be accounted for explicitly. All the nonstructural coefficients can be changed, as desired, to evaluate the sensitivity of projected changes. For example, the manpower per dollar of output pricing coefficients could be changed as one way to study projected changes due to inflation. This, in fact, has already been tested in the initial prototypes of the multilevel model, which contained only a civilian submodel. These pricing coefficients can also be changed when improvements supply better information than is obtainable from historical experience only.

Other model structures might prove to be more worthwhile within the various submodels. This is particularly true of the military submodel where such factors as planned rotation dates and sea/shore assignments should be considered if possible. For example, the military assignment model given in the next section of this paper may be applicable. Additional thought might also be given to an additional modeling or computer-assisted subsystem to aid in the determination of the manpower requirements data.* It might also be desirable to account for risk or probabilistic variations in the requirements data. This has been discussed in terms of the use of chance-constrained programming in [4.4] and [4.5] including the possible implications for the PPB process.

In [4.1] prototype results were given for a multilevel model which includes historical input-output data from the Navy Requirements Model (NARM) and civilian manpower data from the Computer-Assisted Manpower Analyses System (CAMAS) developed by OCMM. Prototypes have been completed of multiactivity linked models (i.e., several manpower models linked by a common ceiling) and of multiprogram linked models. The structure of these models is given in Figure 5. These prototypes involved the Naval Facilities Engineering Command (NAVFAC) and the Naval Underwater Systems Center (NUSC), Newport, R. I., respectively.† Those prototypes, which were generally small scale (i.e., 1,000 equations or less), have been helpful in testing CAMAS with operational data. A larger-scale (3,200 equations and 5,200 variables) model was successfully run in 1971. This involved a civilian

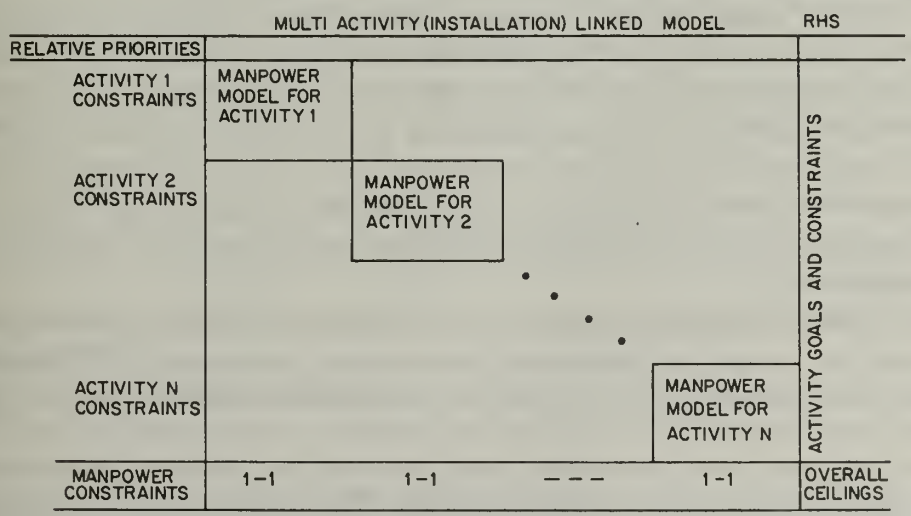


FIGURE 5. Multiactivity (installation) linked model.

*This possibility is already being addressed in a prototype of a civilian manpower model we have under development with the Naval Underwater Systems Center, Newport, R.I.

recruiting requirements model covering the white collar occupations of the Navy work force. This larger model ran in 16 min on a CDC 6600 computer and gave results which were sufficient to encourage movement toward operational implementation.

Some 85 operational data studies have been supported by the use of the CAMAS system. These studies for the most part have been concentrated in the areas of average-grade policy testing and civilian career management planning. However, once the NAVFAC command level studies and the NUSC installation level studies are completed, a comprehensive implementation plan is to be developed and promulgated. This again will be a phased plan in which basic and applied research are an integral part of the implementation process.

INTERFACES WITH DYNAMIC ASSIGNMENT MODELS

We now focus on some potential Navy applications for these models. There are a number of ways the outputs from the global multilevel model might be interfaced with assignment models for operational planning.* For instance, one might consider the assignment models as part of a multistage process. Here one could solve the global model and then use the results to define a space of possible solutions for the more detailed assignment model. That is, the results from the aggregative model could specify the goals for the detailed model. The goal programming format could accommodate discrepancies which would be minimized. Such assignment processes could be extended via dynamic extensions of the classical (static) assignment model, and, in fact, one could then require that the first period on-board projections of the global model be met precisely.

A number of assignment models could be used to make these extensions to the classical model. Here, however, we will use the multiattribute extensions proposed in [4.6] for organization design analyses and restrict ourselves to a somewhat more aggregate model. The latter can be developed for applications to military planning via a simple goal programming model that can be used to coordinate assignment officers and placement officers (over time, if one wishes).

In this model we assume that each assignment officer, being responsible for certain particular men, knows how many of each "type" (e.g., designator and rank) he needs to assign in the current period. Similarly, we assume that each placement officer knows what vacancies he has for the various "types" during the current period due to (a) completion of tours of duty, (b) remaining unfilled billets, and (c) newly authorized billets. We can now choose our goals to come as close as possible to fulfilling each "billet-type" requirement of the various placement officers. Also we can weight the goal-discrepancies according to relative priorities for filling these requirements and require that we do not overfill any requirement.

We can represent this goal programming model for the current period (and similarly for multiperiod situations) in an equivalent "distribution" model format. The structure of such a static assignment model is shown in Figure 6. It should also be noted that an additional column and an additional

---

*There have been several Monte Carlo simulation systems which have been developed for this purpose. A successful U.S. Navy system includes the ADSTAP system which was developed under the direction of R. Lehto of BUPERS for enlisted personnel planning. For a report on ADSTAP see Silverman [14]. The analytical models reported in this present paper differ from these Monte Carlo systems in that it is possible to obtain a closed solution which simultaneously considers all of the variables. Also, it is generally possible to obtain faster computer processing times with less computer memory since aggregates rather than individual records are processed during the simulation phase.

† See A. Charnes, W. W. Cooper, R. J. Niehaus and K. Padalino, "Integrated Manpower and Program Planning Models for Laboratory Management," Civilian Manpower Management, VIII, No. 2, Summer 1974.

	Activity 1			Activity 2			Activity 3			Inventory
	ENS	LTJG	LT	ENS	LTJG	LT	ENS	LTJG	LT	
ENS	1[a] 110	3	10	5 10	2	10	1 70	2	10 10	200
LTJG	20	1 40	3	20	1 50	3	10	1 30	2 10	130
LT	30	25	1 25	20	15	1 15	30	25	1 20	60
REQUIREMENTS	110	40	25	10	50	15	70	30	40	

[a] Relative priorities are indicated by numbers in the upper left hand corner and the other number represents the officer assignments. No entry in the latter position means zero are assigned.

FIGURE 6. Static classical assignment model officer example (with relative priorities as goal weights).

row could be added. See Chap. XIV in [3]. The additional column would provide for the case where we have more personnel than billets; the additional row would provide for the case where we have more billets than personnel. Since this is a distribution model very large and very fast computer algorithms are available.* Thus, the main problem of testing and implementation is one of obtaining and validating the required data inputs in ways that are (or can be made) operational. Along these lines, a prototype test using operational data in approximately 200 equations has been completed by LCDR Cass of BUPERS who is working closely with the OCMM research team.

As mentioned earlier, the static model can be extended to a multiperiod model by using estimates of officer availabilities and billets for succeeding periods. The resulting multiperiod model would then have the same mathematical structure as the single period model, but would be much larger.

Rather than carrying this development further, we have developed a new dynamic model which includes planned rotation dates. Our new dynamic model includes the use of what is called a "multipage"* goal programming model. In this model we consider as a "page" the "generalized network" which we associate with each set of periods associated by transfer within our planning horizon. We then consider transfers of "men" from "activity-billet" positions to other "activity-billet" positions taking into account rotation, or tour of duty, and effects of promotions and retirements. This linkage between the pages or generalized networks is accomplished by means of input-output conditions specified by Markov or manpower transition matrices. The objective of this dynamic model is to minimize weighted discrepancies, across periods, from the authorized billets while considering the constraints imposed by current assignments and planned rotation dates. A formal mathematical statement of this model is included as appendix A.

Various simplifications and approximations of this dynamic model are possible for computational feasibility. More importantly, one has an analytic tool with all the evaluational possibilities of linear programming available.

The interfaces we have discussed have been concerned with models which use some aggregation of the man-job combinations. This discussion could also be extended to include the static and dynamic

* See especially the recent work by F. Glover, D. Karney, D. Klingman and A. Napier [7]. See also V. Srinivasan and G. L. Thompson [15].

* See Chap. XX in [3].

multiattribute assignment models although then we would have had to explode the problem to include the individual job attributes. It is at this level of detail that dynamic mixed manpower systems might be constructed for use in areas such as equal employment opportunity planning and local organization design studies.

MULTIATTRIBUTE ASSIGNMENT MODELS

The multiattribute assignment models operate on the individual man-job matching level. An example of the data used in checking out a toy problem of the static model is given in Figure 7. The occupational interest data in this case involve the use of the Kuder Occupational Interest Blank (KOIB). These are grouped in one set of columns for each of the tabulations—Job Data and Person Data—in Figure 7. The work force profile data grouped in the other set of columns are from the Labor Department's General Aptitude Test Battery (GATB). The solution to this toy problem, which is given in [8], is not the main point of interest here. We are directing attention rather to the multiattribute character of both jobs and persons which are to be considered. This illustrative example involves a model which is static, but it does include many attributes which can have desired values in multiattribute form which includes both goals and lower bounds. Decisions to hire, train, deskill jobs, re-

JOB DATA

Attribute / Job	KOIB interest data		GATB work force data		
	Cler.	Comp.	Verbal	Num.	Cler.
Accounting Clerk	7[a] 6	8 6	x 80	x 100	x 95
File Clerk	8 6	4 0	80 75	75 60	x 105
Stenographer	8 6	5 3	125 105	x 95	x 110
Receptionist	7 5	4 2	125 105	x x	115 105
Supervisor	8 6	8 7	125 105	125 100	115 105

[a] Top values are desired or goal values, bottom values are lower bounds.
x = no value specified.

PERSON DATA

Attribute / Person	KOIB interest data		GATB work force data		
	Cler.	Comp.	Verbal	Num.	Cler.
J. Doe	9	7	130	132	122
Jane Roe	7	5	80	80	110
Harriet McG.	5	7	90	115	95
Dianne Z.	8	6	125	130	119
Sally G.	7	7	115	115	110
Arlene D.	6	7	120	110	110

FIGURE 7. Static multiattribute assignment model.

combine or enrich jobs, and transfer personnel can all be evaluated by such a model. Thus, even though the illustrative example is static it already exhibits capabilities for dealing with more attributes and for exploring more person-job combinations under a greater variety of conditions which greatly exceed those currently being employed in personnel management.

An important consideration in both static and dynamic multiattribute assignment models is to decompose organizational needs into job elements. Job elements may be synthesized from expected or known tasks, skills, and knowledge. To some degree, tasks and skills are already described by occupational codes such as those developed by the U.S. Civil Service Commission.* Evidently these will not be adequate for the applications we intend, however, and so B. Moore et al., are studying how we might best advance job element descriptions to a new level of facility and measurement. The underlying assumption is that job elements can be determined to define absolute levels of education, technical training, and skill with sufficient validity to allow explicit consideration of how job elements might be combined or recombined performing organization tasks or for achieving organizational goals. When included in a multiattribute assignment model these possibilities would then provide a considerable enhancement in evaluating training, upgrading, transfer, and deselection decisions.

Under normal circumstances civilian jobs are filled one-at-a-time. In this situation a simple rank ordering of candidates may be sufficient. However, static assignment models may be useful in the case of apprentice training programs where a number of graduates need to be placed simultaneously. Another case might be in large activities which hire a number of similar people at the same time—e.g., a naval laboratory might hire a number of trainee scientists and engineers at the end of the school year. A similar situation might exist in across-command management intern programs such as the Professional Development Center (PDC) program of the Naval Facilities Engineering Command. The static assignment models may also be useful in testing alternatives during large-scale organizational changes.

The static assignment models appear to have a much wider application to military personnel management systems. One early application of the static multiattribute assignment models might be with the officer distribution planning at BUPERS. In this case there are many instances where a number of personnel must be detailed during a relatively short time period. Similar uses might be made for enlisted personnel. Here a version of the static model could be used for negotiation of first assignments with potential enlistees.

Of course, such static models also invite exploration of their possible dynamic extensions. This, too, is being assayed both analytically and in principle while staying close to experience and the potentials for developing the needed insights in behavioral-science as well as management science modeling possibilities. As a matter of principle, however, dynamic multiattribute assignment models should find many uses in both the military and civilian manpower systems. In the military case these models are especially suited to the combinations of training and experience that the related tours of duty imply for future assignments. In the civilian manpower systems the dynamic models are suited to civilian career planning at the local level and to equal employment opportunity planning. The dynamic models should also be quite useful in planning for filling those jobs which might be filled by either military or civilian personnel.

Development of the requisite materials (including assessment of currently available research) is

* See also U.S. Department of Labor *Dictionary of Occupational Titles* (3rd edition, 1965).

already underway, and exploration of potential applications is planned to obtain validation of such applications of dynamic multiattribute models. We also want to extend the dynamic models to take account of person-to-person as well as person-job interactions. Models of this kind will be needed for the kinds of dynamic organization-person-task designs that is an ultimate objective in this research.

We also want to take account of career aspirations and possibilities across jobs (and periods) as well as within them; this then leads us to organizational and career planning aspects of this research such as is set forth in our spectral analyses model [4.7]. To see what is involved we may observe that it is not adequate to map out only one career possibility. Rather a spectra of careers in an organization will need to be considered perhaps in terms of their probabilities or probability mixtures (i.e. spectra). The structure of an activity or installation level modelling system in which this model could be used is given in Figure 8. Although it cannot portray the dynamic system adequately, this at least combines the areas of program manpower requirements, current career paths, and employee career aspirations. The underlying objective of the system would, in any event, be the minimization of the differences between (a) organizational goals, (b) current manpower trends, and (c) employee aspirations.

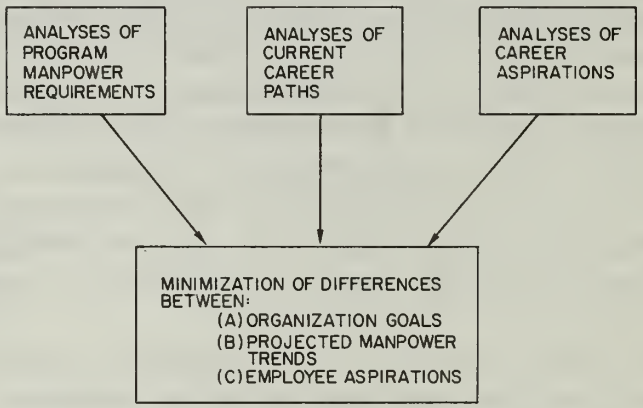

ANALYSES OF PROGRAM MANPOWER REQUIREMENTS
ANALYSES OF CURRENT CAREER PATHS
ANALYSES OF CAREER ASPIRATIONS
MINIMIZATION OF DIFFERENCES BETWEEN:
(A) ORGANIZATION GOALS
(B) PROJECTED MANPOWER TRENDS
(C) EMPLOYEE ASPIRATIONS

FIGURE 8. Dynamic activity manpower modeling system.

Research into such uses of spectral analyses model will be phased into the OCMM research program, however, only after experience has been gained in the applications of the static and dynamic multiattribute assignment models. This seems prudent since these models represent a considerable extension of current practice and therefore involve problems of relevance and communication as well as validation.

SUMMARY

This report provided a review of the wide range of manpower modeling research which we have underway at OCMM, at the University of Texas (Cybernetics Research Center), Carnegie-Mellon University (Management Science Research Group), as well as in other parts of the Navy such as BUPERS and the Naval Underwater Systems Center. This paper also tried to provide a conceptual scheme for this research including a delineation of the ultimate as well as immediate uses of its products. The discussion was pointed toward a system of models for use in managing military-civilian manpower interactions as one example of a mixed-manpower system. For this, we indicated how we had extended previously developed multilevel models to include in one management system: (a) program

planning, (b) civilian manpower planning, and (c) military manpower planning. This phase of the research, which is in addition to the earlier OCMM models now being used in the Navy, has proceeded to a stage in which one global model has already been tested with small prototypes involving program planning and civilian manpower planning data from existing information systems. The inclusion of military data for at least one model structure appears to be an extension which involves more complexity due to size rather than in model technology. But this is not all; it also opens other longer range possibilities, such as we discussed in terms of new goal programming models of the static and multipage varieties. Note, in particular, that this includes their possible integration with the global models.

As a still more remote, but possibily even more valuable extension we discussed static and dynamic multiattribute assignment models. This included implementation possibilities (due to the development of a new solution algorithm for the static multiattribute models which allow transformation to an ordinary assignment model structure) which must also be accompanied by suitable behavioral science research and guidance from management use and experience. The combination of career spectra and organization design in relation to organization goals and behavior also offers exciting possibilities. Those will be developed in future reports as research results justify their release. For the present, however, the final chapter [4.1] can provide further insight into some of those possibilities and what is required for their attainment.

BIBLIOGRAPHY

[1] Augusta, J., "Navy Resource Study Progress Report and Proposed Program," Center for Naval Analyses (INS) 1188–72 (15 Aug. 1972).

[2] Bagley, D. H., "The Future of the Navy, Military and Civilian, Personnel Structure," The Journal of Navy Civilian Manpower Management, Vol. VI, No. 3 (Winter 1972).

[3] Charles, A. and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming (New York: John Wiley & Sons, Inc., 1969).*

[4] Charnes, W., W. W. Cooper, and R. J. Niehaus, *Studies in Manpower Planning* (Washington: Office of Civilian Manpower Management, 1972). This volume includes [4.1] through [4.7].

[4.1] Niehaus, R. J., *The Application of Computer-Assisted Multi-Level Manpower Planning Models in the Federal Government*, D. B. A. Dissertation, The George Washington University (Ann Arbor: University Microfilms, 1972).

[4.2] Charnes, A., W. W. Cooper, and R. J. Niehaus, "A Goal Programming Model for Manpower Planning," Chap. 4 in *Management Science in Planning and Control*, John Blood, ed. (New York: Technical Association of The Pulp and Paper Industry, 1968).

[4.3] Charnes, A., W. W. Cooper, R. J. Niehaus, and D. Sholtz, "A Model for Civilian Manpower Management in the U.S. Navy," in *Models of Manpower Systems*, A. R. Smith, ed. (London: The English Universities Press, 1970).

[4.4] Charnes, A., W. W. Cooper, R. J. Niehaus, "A Generalized Network Model for Training and Recruiting in Manpower Planning," in *Manpower and Management Science*, A. R. Smith, ed. (London: The English Universities Press, 1972).

[4.5] Charnes, A., W. W. Cooper, R. J. Niehaus, and D. Sholtz, "Multi-Level Models for Career Management and Resource Planning," in *Proceeding of a 1971 NATO Conference on Man-Power Planning Models* D. Clough, (C. C. Lewis and A. L. Oliver, eds., London, English Universities Press, 1974).

[4.6] Charnes, A., W. W. Cooper, R. J. Niehaus, and A. Stedry, "Static and Dynamic Assignment Models with Multiple Objectives and Some Remarks on Organization Design," *Management Science, 15* (Apr. 1969).

[4.7] Charnes, A., W. W. Cooper, R. J. Niehaus, and D. Sholtz, "An Algorithm for Multi-Attribute Assignment Models and Spectral Analyses for Dynamic Organization Design," presented at 41st Operations Research Society of America Meeting, New Orleans, Louisiana (27 Apr. 1972).

[5] Charnes, A., W. W. Cooper, D. Klingman, and R. J. Niehaus, "Static and Dynamic Biased Quadratic Multi-Attribute Assignment Models: Solutions and equivalents," Research Report CCS115, Center for Cybernetic Studies, University of Texas (Mar. 1973).

[6] Gewirtz, M., S. Korn, and R. J. Niehaus, "Application of Civilian Manpower Models to the Naval Facilities Engineering Command," OCMM Research Report No. 16, U.S. Navy Office of Civilian Manpower Management, Washington, D.C. (Jan. 1974).

[7] Glover, F., D. Karney, D. Klingman, and A. Napier, "A Computation Study on Start Procedures, Basis Change Criteria, and Solution Algorithms for Transportation Problems," Management Science *20* (Jan. 1974).

[8] Larson, D., D. Sholtz, and E. Trimble, "A Numerical Example of a Static Multi-Attribute Assignment Model," OCMM Research Report No. 15, U.S. Navy Office of Civilian Manpower Management, Washington, D.C. (Oct. 1973).

[9] Hibbs, N., "An Introduction to NARM," Center for Naval Analyses Memorandum (CNA) 1684–72 (1972).

[10] Mannis, L., "Application of the Navy Average Grade Model to the Naval Underwater Systems Center," NUSC TM No. MA 11–4536–72.

[11] Niehaus, R. J., D. Sholtz, and G. L. Thompson, "Guessing Your Way to Recruiting Goals," Journal of Navy Civilian Manpower Management, VII, No. 2, (Summer 1973).

[12] Program Definition Study for the Navy Manpower Planning System, Orkand Corp. Report No. TR–71W–001, BUPERS Contract No. N00–032–71C–0024.

[13] Ruefli, T. W., "A Generalized Goal Decomposition Model," Management Science *17*, B–5D5–518 (Apr. 1971)

[14] Silverman, J., "Personnel Resource Accounting in an Operational Environment," in D. Clough, C. C. Lewis and A. L. Oliver, eds, *Manpower Planning Models*, (London, English Universities Press, 1974).

[15] Srinivasan, V. and G. L. Thompson, "Alternative Formulations for Static Multi-Attribute Assignment Models," Management Science *20*, (Aug. 1973).

[16] Weitzman, R. A., ed. *Proceedings of a Naval Conference on Naval Research:* with special reference to the All-Volunteer Force (Monterey, Calif.: Naval Postgraduate School, Oct. 1972).

INSURANCE TYPE ITEM PROVISIONING GUIDELINES*

Robert W. Burton

Thomson and McKinnon, Auchincloss, Kohlmeyer, Inc.
Palo Alto, California

and

Stratton C. Jaquette

Management Science Associates
Los Altos, California

ABSTRACT

Policy decisions for insurance type items, where zero or one unit is maintained at the depot, are more difficult and more critical than decisions for common supply items. This report presents results of developing initial provisioning guidelines for insurance type items. The guidelines are based on examination of lifetime costs and benefits. Costs of stocking an item as compared with not stocking are developed through a sinking fund annual payment formulation. Benefits of stocking are developed as stationary reduction in time weighted backorders experienced. A resource allocation formulation yields an optimal policy for allocating a fixed budget. The guideline is presented with refinements based on a sample of items. A figure of merit is calculated for each item, and if it is large the item is stocked while if small it is not stocked. Empirical definitions for large and small are developed based on sample data. Estimation techniques are discussed for deriving all of an item's parameters needed to compute its figure of merit. A Bayes procedure is suggested based on family group Experienced Demand Replacement Factors. This and other techniques are discussed.

1. INTRODUCTION

Policy decisions for provisioning of insurance type items in the Navy are more difficult and more critical than decisions for common supply items. This is evident since insurance items are usually expensive items with a low and often sporadic demand history. Overstocking of high demand items, although it may be expensive in terms of high cost of current inventory, is self-correcting in a relatively short time. Overstocking of low demand items is not immediately self-correcting, and the misallocations can last for a much longer time. With low demand items there is also the high probability that the excess stock will never be used and will become obsolete. What makes the policies surrounding insurance items even more critical is that, in initial provisioning, a very large percentage of the items typically fall into this category.

This paper reports on the results of a study undertaken to develop initial provisioning guidelines and operating policies for insurance type items. These results, if implemented, should improve the cost effectiveness of the supply system. Since the emphasis was on implementation the first part of

*This work was supported by the Office of Naval Research and Naval Supply Systems Command under contract N00014-73-C-0003 while the authors were with Science Applications, Inc.

this study involved determining pertinent data that could be made available for estimating an item's characteristics. To employ this data a cost effective provisioning scheme also required developing effective parameter estimation procedures. The devised guidelines were then evaluated and refined in a series of tests on a sample of insurance type provisioning items.

2. BASIS FOR PROVISIONING DECISIONS

For insurance type items, or any item, supply is concerned with minimizing procurement costs while at the same time minimizing backorders over the item's lifetime. In an attempt to balance increased purchase costs with reduced backorders some reasonable method needs to be available to compare annual costs with annual time weighted backorders over the item's lifetime.

2.1. Comparison of Policy Costs Over Time

The two supply policies considered here are stock and no stock. Each of these policies will generate a sequence of costs over time, that is the item purchase costs that occur when demands occur. Figure 1 shows these costs graphically. The element of interest in this figure is the difference in costs for the two policies; the supply system must incur the no stock policy sequence of costs, but it has the option of incurring the stock policy costs as an alternative.

Ignore, for the moment, costs other than capital costs: holding costs and inventory maintenance costs can easily be incorporated into the interest rate or discount factor. It is evident that the difference in costs between these two policies is the original cost of placing the item in inventory minus the discounted salvage value of this item (if any) at the end of its useful life.

At the time of provisioning the current worth of an item's salvage would be difficult to determine. In Navy applications, however, this value tends to be insignificant. The gross salvage value depends on (i) what the item will be worth when disposed of (scrap value, resale value, etc.) and (ii) the future point in time at which use of the item will be discontinued. A net salvage value can also be computed by deducting the cost of the salvage operation. Historically, the gross salvage value has been in the order of 5 to 10 percent of cost. Net salvage values have historically been about zero, thus a value of 0.05 of initial cost, C, is a reasonable upper limit as the net salvage value. The discounting, using an interest rate of 15 percent and a lifetime of 10 years, reduces the present worth of the salvage to less than 0.005 C. Ignoring this small amount should not affect the cost estimates used in provisioning policies.

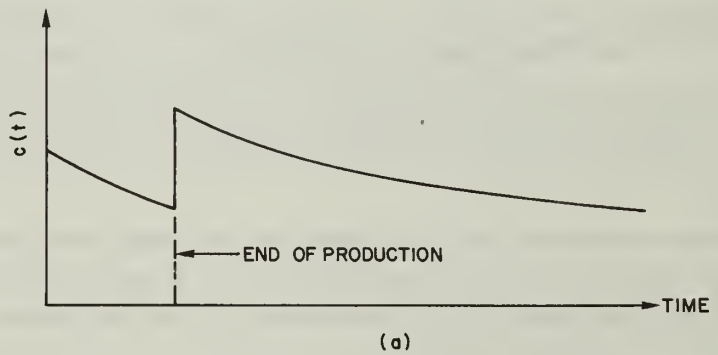


FIGURE 1. (a) Plot of present value of a unit purchased at time t.

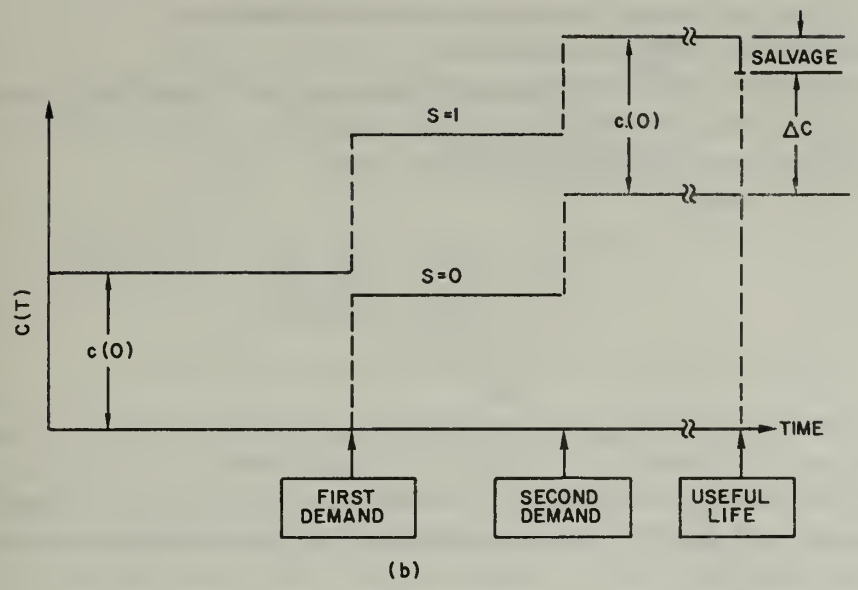


FIGURE 1. (b) Present cost as a function of time for a given demand history.

How should the difference in cost of two policies be compared? To compare the stock to the no stock policy, one might say that the difference in cost is the current cost of buying one item. This procedure, however, does not consider the lifetime consequences of the two policies. To illustrate this point consider two insurance type items of equal cost and with the same estimated demand rate. Assume, however, that one is expected to be purchased for 5 years and the other for 10 years. The immediate out-of-pocket cost of having a stock policy for either item is the same, but one will have a beneficial effect on backorders for 5 years while the other will have a beneficial effect for 10 years. Clearly, the item with the longer expected usage is preferred for the stock policy. Thus the useful life, or time to obsolescence, should play a role in insurance item stocking decisions.

2.2. Sinking Fund Cost of Provisioning

When a business facility is considering a capital expenditure, say for a plant or a piece of equipment, it usually weighs the return to be gained from the expenditure with the costs incurred. A profitable way of funding the expenditure is to borrow the dollars and to pay this loan back using a portion of the return generated by the capital investment. A convenient way of doing this is by setting up a sinking fund of equal payments over the lifetime of the plant or equipment. The desirability of the capital expenditure can then be determined by comparing the annual return to its annual cost, i.e., the sinking fund payment.

This approach can also be taken to compare a no stock to a stock policy for insurance items. The cost used in selecting the policy is not the cost of the unit, but rather the annual sinking fund cost computed over the expected life of the item. Thus the cost is judged on an annual basis rather than its immediate out-of-pocket cost. This sinking fund cost thus generates an annual cost of provisioning for the item.

The expression for the sinking fund payment is easy to derive by setting the amount borrowed equal to the present worth of the sinking fund payments. If

P is the sinking fund payment (made at the beginning of each year),
I is the interest rate (time value of money),
N is the time during which the FSN may be bought, i.e., time to replacement obsolescence,
C is the amount borrowed (current cost of the item), and
β is the discount factor defined as $\beta=(1+I)^{-1}$, then

$$P=\frac{C\left(1-\frac{1}{1+I}\right)}{\left[1-\left(\frac{1}{1+I}\right)^{N}\right]}=\frac{C(1-\beta)}{(1-\beta^{N})}.$$

In this equation the factor $(1-\beta^N)/(1-\beta)$ indicates what fraction of C must be paid each year into the sinking fund. For example, if $(1-\beta^N)/(1-\beta)$ equals 10 then the annual cost is one-tenth of C. The behavior of this factor for various values of N and β is illustrated in Figure 2. Note for the smaller values of β (high interest rate) it is seen that the sinking fund payments are rather insensitive to N beyond a very few years.

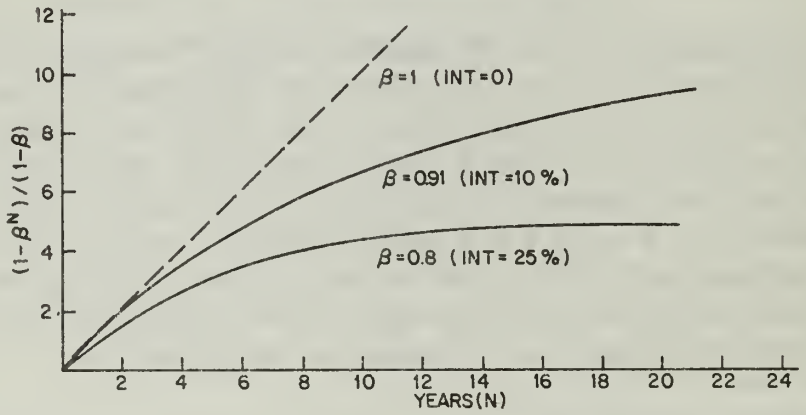


FIGURE 2. Sensitivity of sinking fund payments to lifetime (N) and discount factor (β).

2.3. Supply System Performance Resulting from Provisioning

If a fixed inventory policy is assumed, i.e., always stock one unit or never stock, the backorders experienced over the lifetime will be expected to be constant from year to year. From the result given in Burton and Jaquette [2], instantaneous backorders in the steady-state case are expected to be λT if the item is not stocked and $\lambda T-(1-e^{-\lambda T})$ if the item is stocked, where λ represents the items demand rate and T the expected resupply time. Thus the benefit of the stock policy over the no stock policy is $(1-e^{-\lambda T})$ backorder-years in each year of the item's life (see Figure 3).

It is clear, then, that the decision to stock, if carried out for the entire lifetime, is expected to provide a steady stream of benefits equal to $(1-e^{-\lambda T})$ backorder-years per year for the lifetime of the item. The benefit of $(1-e^{-\lambda T})$ can then be compared with the equivalent "sinking fund" annual cost of $(1-\beta)C/(1-\beta^N)$. The decision to stock or not can be made on the basis of the value of the annual benefit as compared with the annual cost.

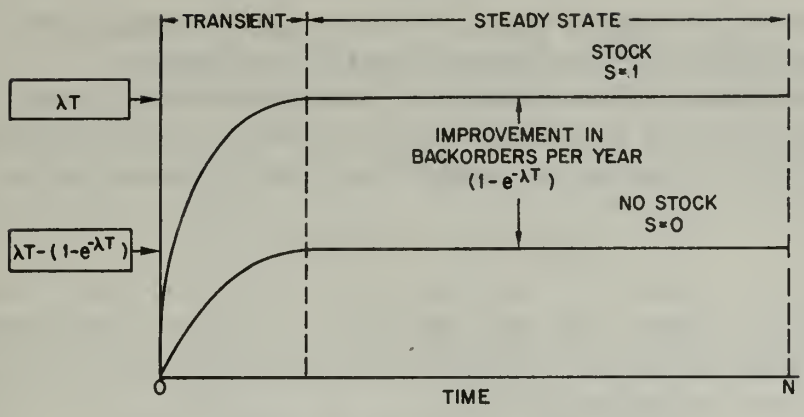


FIGURE 3. Time weighted backorders over an item's lifetime.

This result, of course, depends on the validity of the steady-state assumptions. However, from the results of the appendix it is clear that the probability transient effects on the number of backorders at any point in time are significant only for a very short time period. The transient period for backorders, typically 6 months or less, is generally very short compared with the lifetime of most items. For instance, if a lifetime of 8 years is estimated, the total backorder days accumulated over the lifetime ignoring the transient will be at most a few percent over that including the transient.

3. STOCKING CRITERIA BASED ON INITIAL PROVISIONING FUNDS

The ultimate objective of this paper is to provide a cost effective method for distributing the initial provisioning budget which is not only reasonable, but also justifiable. The procedure which is suggested is to make this decision based on a figure of merit calculated for each item. If this figure of merit is large, stock the item. We define large and small in section 4.

In previous work, Burton and Jaquette [2], the resource allocation problem is posed in classical form:

$$\text{minimize} \sum_{items} \text{(backorder-years expected per year)}$$

$$\text{subject to} \sum_{items} \text{(this year's initial provisioning cost)} \leq \text{(this year's budget).}$$

This is a reasonable formulation and quite adequate for many purposes. The one disadvantage it has, however, is that it is very myopic. Consideration is only given to the initial cost for stocking items, since only this year's budget is considered. Item lifetimes and expected initial provisioning costs per year are ignored.

Long run or total costs can be reduced if the backorders and costs over the item's entire lifetime are considered. It is better to allocate this year's funds on the basis of a complete lifecycle objective than to view the problem myopically and thus follow a suboptimal policy. The following resource allocation formulation does this:

$$\text{minimize} \sum_{items} \text{(backorder-years expected per year)}$$

$$\text{subject to (annual provisioning cost)} \leq \text{(fixed annual budget)}$$

This new formulation, although only slightly different from the one posed earlier, allows simple and direct accounting of lifetime costs and backorders over an item's lifetime.

The results obtained Burton and Jaquette [2] concerning resource allocation is applicable to the present problem. Define a measure of an item's desirability for stocking by the following ratio of benefit to cost which is derived as a dual variable from the above allocation problem:

$$\gamma = [(1-\beta^N)(1-e^{-\lambda T})]/[(1-\beta)C]. \tag{1}$$

An optimal provisioning policy can be easily obtained if a γ^* is found such that purchasing all items (considered in a fiscal year) whose γ value is greater than or equal to γ^* exactly expended the fiscal budget. The optimal policy would be to

(i) calculate the figure of merit (1) for each item,
(ii) stock all items with $\gamma \geq \gamma^*$, and
(iii) select γ^* such that the cost of provisioning just equals the budget.

In actual applications the above procedure appears overly strict; such a γ^* would be difficult to calculate. It would also fluctuate with the budget, and would depend on what items were to be considered during the year, their costs, and their other characteristics. Even if these factors were known, there would still be considerable debate about items with $\gamma \doteq \gamma^*$. If $\gamma \doteq \gamma^*$, the provisioning decision would be based on such questions as essentiality or data errors.

It is thus advisable to use a procedure that is easier to use. The procedure which was developed was to set two limits $\bar{\gamma}$ and $\underline{\gamma}$ and to stock an item if its $\gamma \geq \bar{\gamma}$ and not to stock it if $\gamma \leq \underline{\gamma}$. The decision for items whose γ lies between $\underline{\gamma}$ and $\bar{\gamma}$ is left to further investigation. This procedure should be very simple to implement, as the calculations and data requirements are relatively simple. (It should be clear that the γ^* mentioned above will lie between $\bar{\gamma}$ and $\underline{\gamma}$.)

4. PRELIMINARY PROVISIONING GUIDELINE INVESTIGATION

The suggested provisioning guideline was tested and refined using available data on a sample of typical insurance items. The sample consisted of 57 items on which data was complete enough to calculate the provisioning criterion function (1) with some confidence.

4.1. Sample Data

A sample of over 57 items was selected in coöperation with the Ships Parts Control Center (SPCC) for use in testing the provisioning guideline. These items were chosen so that roughly half would show experienced demand rates between one and four a year and half would show demand rates less than one a year. (Obviously, this sample is not strictly representative of all insurance type items although it proved extremely useful in these tests.) On each item an obsolescence rate, a demand rate, demand history, the procurement lead time, and the unit replacement price were obtained in order to demonstrate the guideline. This data was obtained from the Master Data File (MDF). As expected, the obsolescence rates in the MDF were arbitrary. Rates of 0.33 or 0.10 were indicated for all items. These rates were therefore not used in the test analysis. The sample of 57 items is given in Table 1.

TABLE 1. *Provisioning Guideline—Rank by Using Beta = 0.909, Lifetimes of 10 Years*

Rank	PCILE	Name	FIIN	PROCLT	Price	OBSOL	DMSY	DMDOB	Life	Gamma	Prov Cost
1	0.9	Arm, throttle plate	8721500	1.80	$0.84	0.33	0.363	1.318	3	3.8600959	$0.84
2	2.6	Shaft, pump, driven i	3349431	2.70	0.97	0.33	0.222	0.182	3	3.1416150	1.81
3	4.4	Vane, compressor sta	8957055	1.70	0.88	0.33	0.225	0.545	3	2.4412834	2.69
4	6.1	Spring, helical, comp	8234303	2.70	4.31	0.33	0.778	0.364	3	1.3762957	7.00
5	7.9	Spring, helical, comp	3164153	2.70	0.75	0.33	0	0.045	3	1.0310571	7.75
6	9.6	Cover, crankcase low	3371998	2.70	5.90	0.33	0.294	0.273	3	0.6276459	13.65
7	11.4	Connection, supercha	3329613	2.80	12.00	0.33	0	0.818	3	0.5062363	25.65
8	13.2	Bushing, dowel, taper	3432823	2.70	4.00	0.33	0.095	0.227	3	0.3822988	29.65
9	14.9	Thermostat, flow con	8612294	1.70	1.55	0.33	0.045	0.136	3	0.3211499	31.20
10	16.7	Guide, control	3377410	3.70	17.31	0.33	0.439	1.000	3	0.3135275	48.51
11	18.4	Diaphragm assy, turb	8639510	3.90	3.87	0.33	0.050	0.045	3	0.3094224	52.38
12	20.2	Elbow	365484	2.70	2.75	0.33	0.040	0.455	3	0.2516136	55.13
13	21.9	Gearshift, spur	3183713	2.70	15.00	0.33	0.236	0.409	3	0.2123366	70.13
14	23.7	Thermostat, flow con	8222116	1.70	1.76	0.33	0.022	0.091	3	0.1409766	71.89
15	25.4	Housing assembly	2186936	2.70	26.50	0.33	0.202	0.136	3	0.1072237	98.39
16	27.2	Strainer element	3372018	2.70	49.00	0.33	0	0.545	3	0.1062710	147.39
17	28.9	Gear, spur	3247819	2.70	55.00	0.33	0	0.636	3	0.1008240	202.39
18	30.7	Thermostat, flow con	8778326	.90	22.14	0.33	0.384	0.091	3	0.0892053	224.53
19	32.5	Gear, helical	8452567	4.20	55.00	0.33	0.300	0.318	3	0.0880327	279.53
20	34.2	Cover assy	3371999	2.70	63.77	0.33	0.628	0.864	3	0.0865425	343.30
21	36.0	Bracket assy, revers	3183672	2.80	75.87	0.33	0.815	0.682	3	0.0799929	419.17
22	37.7	Gear assy, spur	3247813	2.70	72.00	0.33	0.348	0.273	3	0.0571903	491.17
23	39.5	Shaft, pump driving	3349435	2.70	80.38	0.33	0.378	3.636	3	0.0537847	571.55
24	41.2	Sleeve, reverse gear	3466233	4.60	125.72	0.33	0.510	0.273	3	0.0486147	697.27
25	43.0	Bolt, externally rel	2060434	2.70	136.00	0.33	0	0.773	3	0.0435338	833.27
26	44.7	Screw, externally rc	2063958	2.70	20.17	0.33	0	0.045	3	0.0383388	853.44
27	46.5	Liner, bearing	3377416	2.70	146.85	0.33	0.568	0.545	3	0.0360961	1000.29
28	48.2	Connection assy, E, H	3304633	2.40	156.60	0.33	0.694	0.545	3	0.0350003	1156.89
29	50.0	Adapter assy, oil pu	3247793	2.70	37.50	0.33	0.072	0.227	3	0.0318434	1194.39
30	51.8	Impeller, salt water	3183688	2.10	72.21	0.10	0.179	3.182	10	0.0293283	1266.60
31	53.5	Coupling	8742301	1.30	125.00	0.33	0.504	1.818	3	0.0259904	1391.60
32	55.3	Housing assembly, oi	2186947	2.70	53.00	0.33	0.073	0.136	3	0.0228138	1444.60
33	57.0	Body assy, control v	3377402	5.80	168.00	0.33	0.110	0.045	3	0.0189756	1612.60
34	58.8	Gear assy	3062003	3.20	233.97	0.33	0.299	0.273	3	0.0177918	1846.57
35	60.5	Tool, cylinder barre	3377428	2.20	381.63	0.10	0	1.182	10	0.0163960	2228.20
36	62.3	Impeller	365452	3.20	71.95	0.33	0.059	0.045	3	0.0161623	2300.15
37	64.0	Cutout assy, overspe	3329616	4.40	347.00	0.33	0.396	0.455	3	0.0160678	2647.15
38	65.8	Detent and shaft ass	3331513	2.70	292.00	0.33	0.302	0.045	3	0.0129055	2939.15
39	67.5	Adapter assy	3247802	2.20	231.96	0.33	0.260	0.091	3	0.0126930	3171.11
40	69.3	Stem, test valve	8580097	2.60	65.00	0.33	0	0.045	3	0.0114815	3236.11
41	71.1	Gearshaft, spur	3183687	2.70	562.00	0.33	0.579	0.227	3	0.0095078	3798.11
42	72.8	Damper assy, vibrati	3331523	2.80	637.00	0.33	0.410	1.136	3	0.0072442	4435.11
43	74.6	Shaft, propulsion, sh	8410450	2.50	619.54	0.33	0.431	1.000	3	0.0071956	5054.65
44	76.3	Gear sleeve assy	3247798	3.00	518.00	0.33	0	0.227	3	0.0064444	5572.65
45	78.1	Sprag assy	3371989	2.70	148.60	0.33	0.055	0.136	3	0.0062769	5721.25
46	79.8	Spool and sleeve ass	9101689	3.70	609.00	0.33	0.224	0.318	3	0.0062532	6330.25
47	81.6	Housing assembly, en	2186941	3.10	232.00	0.33	0.064	0.273	3	0.0052429	6562.25
48	83.3	Housing assembly	532659	5.30	986.00	0.33	0	0.091	3	0.0026230	7548.25
49	85.1	Rotor assembly, pump	8769224	1.50	200.00	0.33	0.034	0.091	3	0.0016803	7748.25
50	86.8	Rotor assembly, pump	8590623	1.90	662.00	0.33	0	0.091	3	0.0016211	8410.25
51	88.6	Filter-separator, li	9033136	2.70	1795.00	0.33	0.148	0.136	3	0.0012404	10205.25
52	90.4	Clutch assy	3183718	4.70	4778.00	0.33	0	0.182	3	0.0008132	14983.25
53	92.1	Flywheel and clutch	3899474	6.70	3560.00	0.33	0	0.045	3	0.0004942	18543.25
54	93.9	Filter-separator, li	9033135	2.70	2457.00	0.33	0.068	0.091	3	0.0004614	21000.25
55	95.6	Filter-separator, li	8409851	2.70	2570.00	0.33	0	0.045	3	0.0003009	23570.25
56	97.4	Gear assy, reversc a	3247840	2.70	50060.00	0.33	0	0.091	3	0.0000294	73630.25
57	99.1	Gear assy, reverse a	3247841	2.70	47520.00	0.33	0.017	0.091	3	0.0000064	121150.25

Table 1 summarizes the provisioning criterion function calculations and input data. Columns 3 through 9 indicate the data obtained from the MDF. In this table NAME is the item nomenclature, PROCLT is procurement lead time in quarters, PRICE is the replacement price, OBSOL is the reported obsolescence rate, DMDSY is the MDF system demand forecast, DMDOB is a demand rate derived from historical (past 20 quarters) observations, and LIFE is estimated FSN lifetime. LIFE is given by 1/OBSOL, however a value of 10 years was used arbitrarily for N in all computations.

4.2. Study Results

The results of the applying provisioning guideline to the sample of 57 items are summarized in the following figures. To test the sensitivity of the guideline several computer runs were made varying the items' lifetimes (8 to 15 years), varying the interest (10 to 15 percent), and using MDF demand rate or demand rate based on observations. All results were similar in character and only one series of runs is presented here.

Table 1 presents the sample of 57 items ranked in order of preference for stocking, i.e., ranked on the items' GAMMA (γ) value. Some of the details of deriving this value in this example were (i) a discount rate of 0.909 (a 10-percent rate of interest), and (ii) in cases where the system demand forecast was zero, the demand rate was derived from the observations. The last column, PROV COST, gives the cumulative cost for stocking the items recommended for stocking to that point. For example, the value in row 10 indicates that if the first 10 items were bought, the optimal γ^* cutoff would be about 0.310 and cost would be \$48.51.

It was possible to derive some interesting empirical relationships between the number of items bought, the cost of provisioning, and the gamma cutoff value. Both the inidividual gamma values and the cumulative provisioning cost can be fit reasonably well by the lognormal cumulative distribution function. This result is indicated in Figure 4, which shows the logarithms of the gamma values and normalized cumulative provisioning cost plotted on normal probability paper. Both sets of points can be seen to follow roughly straight lines, indicating that a lognormal function is a good approximation to the data. To avoid needing a scale factor to expand the sample results to a larger population, the cumulative provisioning cost in this example has been divided by 57, the number of items under consideration.

By assuming the sample of 57 items to be representative of insurance type items, the relationship found in Figure 4 can be generalized to any other representative group of insurance items. Thus inferences can be made on the group of items that will be considered for initial provisioning in any given fiscal year. Using this sample the following relationships can be expected to hold:

Let α be the percentage of insurance items considered in a year that are stocked,
B be the provisioning budget for the year,
n be the number of insurance items that will be considered for stocking during the year,
$\Phi(\cdot)$ be the normal cumulative distribution function, and
γ^* be the provisioning criterion cutoff value, i.e., if an item's γ value is greater than γ^*, it will be stocked.

Then the following two independent equations were found to hold:

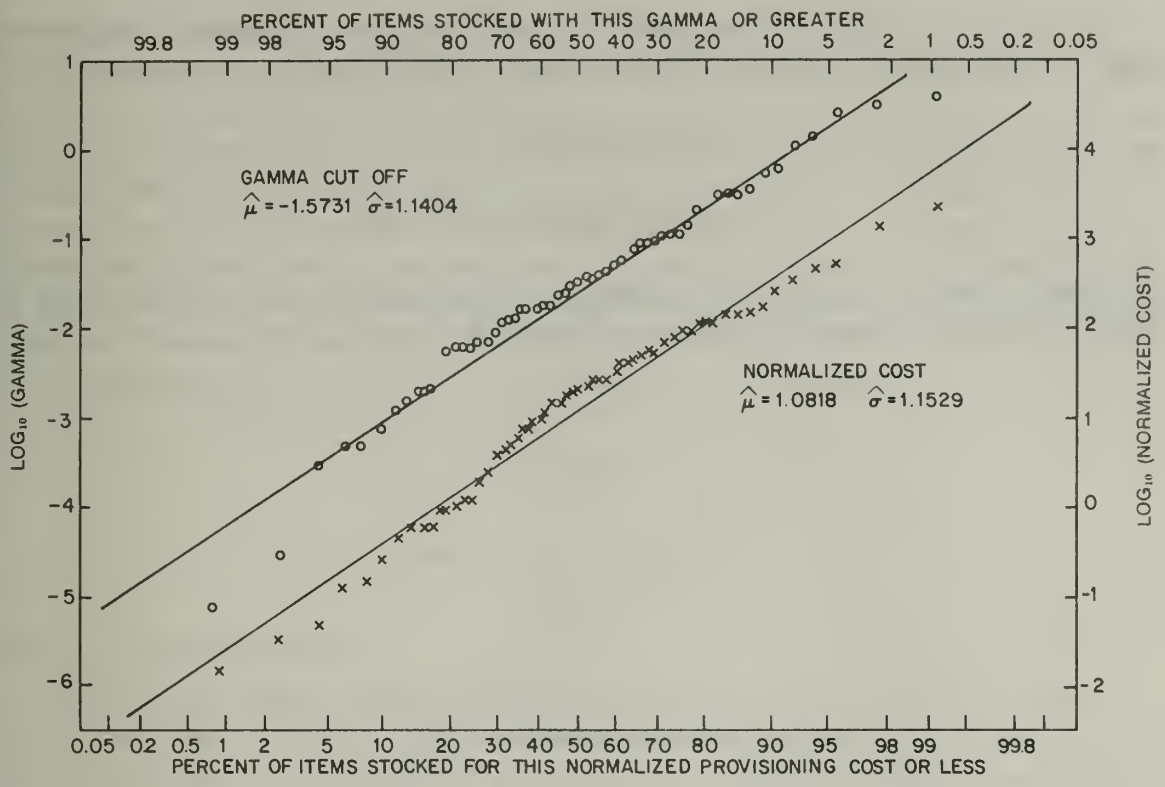

PERCENT OF ITEMS STOCKED WITH THIS GAMMA OR GREATER
99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.05
GAMMA CUT OFF
$\hat{\mu}$ = -1.5731 $\hat{\sigma}$=1.1404
NORMALIZED COST
$\hat{\mu}$ = 1.0818 $\hat{\sigma}$=1.1529
LOG₁₀ (GAMMA)
LOG₁₀ (NORMALIZED COST)
0.05 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8
PERCENT OF ITEMS STOCKED FOR THIS NORMALIZED PROVISIONING COST OR LESS

FIGURE 4. Plots to show log normal relationships between percent of items stocked and gamma cut off and normalized cost.

$$\alpha = 100\Phi\left[\frac{\log_{10}(B/n) - 1.0818}{1.1529}\right]$$

(2)

$$\alpha = 100\Phi\left[\frac{-\log_{10}(\gamma^*) - 1.5731}{1.1408}\right].$$

By using a percentile, α, and normalizing the provisioning cost, these empirical formula can be expected to hold for any representative group of insurance items. The specific constants would have to be recomputed if a different discount rate were used or if different expected lifetimes were available. These relationships indicate that γ^*, the gamma cutoff value for initial provisioning, can be related to (B/n), the normalized provisioning cost/budget figure. Solving equations (2) directly leads to the relationship $\gamma^* = 0.3143(B/n)^{-0.9895}$. An improvement to this relationship can be obtained by performing a linear regression directly on the variables $\log_{10}(\gamma^*)$ and $\log_{10}(B/n)$. This regression analysis results in the following relation:

$$\gamma^* = 0.2794(B/n)^{-0.9422}$$

(3)

$$(R^2 = 0.916 \text{ on logarithmic regression}).$$

This indicates that 91.6 percent of the variation in γ^* can be explained by the function in (3).

The result in (3) can be improved slightly by employing a model of the form $\gamma^* = a + b(B/n)^c$. Using a combination of graphical and analytical techniques, the following result was obtained:

$$\gamma^* = -0.001 + 0.25(B/n)^{-0.78} \tag{4}$$

($R^2 = 0.977$ on logarithmic expression).

This indicates that 97.7 percent of the variation in γ^* can be explained by the function in (4). This relationship is plotted along with the experimental points in Figure 5.

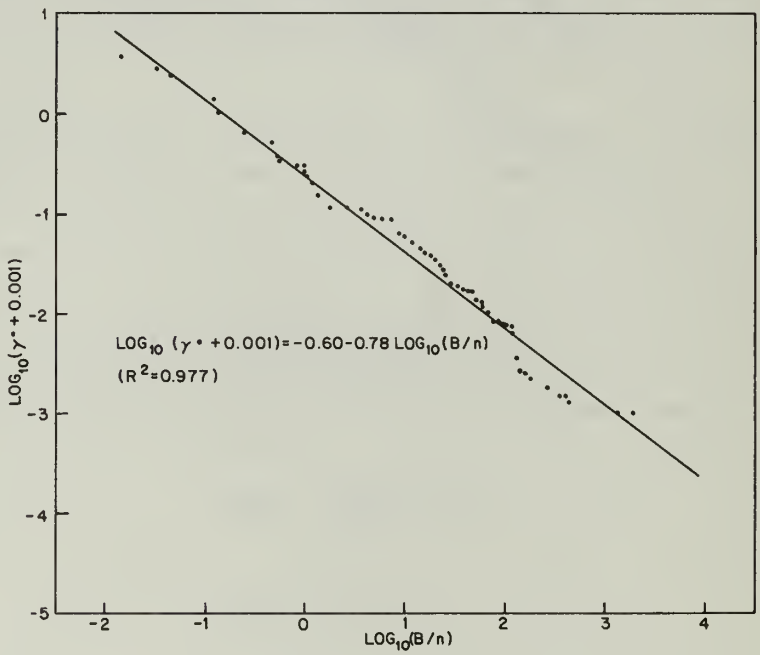


FIGURE 5. Relationship of $\log_{10}$ (gamma cut off +0.001) to $\log_{10}$ (provisioning cost/items considered).

4.3. The Provisioning Guideline

Equations (2) and (3) can be used to determine the specific parameters of the provisioning guideline presented in section 3.2. The first task in implementing this guideline is to set limits, $\bar{\gamma}$ and $\underline{\gamma}$, for the provisioning criterion function. The objective for setting these values is to obtain a working guideline that is insensitive to changes in budgets, data errors, and other unexpected changes. To determine $\bar{\gamma}$ and $\underline{\gamma}$, first estimate the number of insurance items that can be expected to be reviewed for provisioning during the year. Let this be n. Also estimate the maximum and minimum provisioning budgets that can reasonably be expected and let these be $\bar{B}$ and $\underline{B}$, respectively. The first approximation to $\bar{\gamma}$ and $\underline{\gamma}$ can then be obtained using (4) as follows:

$$\bar{\gamma} = -0.001 + 0.25(\underline{B}/n)^{-0.78}$$

$$\underline{\gamma} = -0.001 + 0.25(\bar{B}/n)^{-0.78}. \tag{5}$$

This will insure that budget changes will not void any of the stock or no stock decisions.

Some adjustment in $\bar{\gamma}$ and $\underline{\gamma}$ should be made to avoid an improper provisioning decision due to data errors or poor estimation. The items' γ values (1) are quite sensitive to estimation of λ, T, and C, but not very sensitive to N. The values of $\bar{\gamma}$ and $\underline{\gamma}$ are somewhat sensitive to n. The results of the investigation into estimation methods (section 5) indicates that 5-percent uncertainty in estimation of the demand rate λ can be expected. Assuming similar uncertainty in the other parameters, the γ value calculated by (1) for each item has an uncertainty of ± 15 percent. It is thus desirable that fluctuation of an item's γ value of ± 15 percent should not change the provisioning decision from stock to no stock or vice versa. Therefore $\bar{\gamma}/\underline{\gamma}$ should be 1.15 or greater. If the initial values of $\bar{\gamma}$ and $\underline{\gamma}$ given by (5) are such that $\bar{\gamma}/\underline{\gamma}$ is less than 1.15, then adjusted values should be used as follows:

$$\text{(adjusted) } \bar{\gamma} = \bar{\gamma}[(\bar{\gamma}/\underline{\gamma})/1.15]^{1/2}$$

(6)

$$\text{(adjusted) } \underline{\gamma} = \underline{\gamma}[1.15/(\bar{\gamma}/\underline{\gamma})]^{1/2},$$

where $\bar{\gamma}$ and $\underline{\gamma}$ are as given in (5).

The provisioning decision can then be simplified by using the following guideline. Calculate the γ value for any insurance type item under consideration according to (1). If γ is greater than the $\bar{\gamma}$ given in (6), stock the item. If γ is less than the $\underline{\gamma}$ given in (6), do not stock the item. If γ lies between $\underline{\gamma}$ and $\bar{\gamma}$, the decision should be made after more careful consideration of the item's characteristics and the budget situation.

The simplicity of the guideline makes it attractive as a useful management tool. It is suggested that before implementation the guideline be further refined using a larger sample of items. The additional refinement would be to insure that the parameters used in the guideline are representative of all insurance type items. The impact of varying expected lifetimes should also be determined.

5. PARAMETER ESTIMATION

It is important when implementing any general procedure such as this that accurate and consistent methods be used for estimating all the input parameters of the provisioning criterion function. To use this procedure for consumable insurance type items, and for reparable items if the repair cycle is factored out, the needed input parameters are

N, the expected usable lifetime of the FSN,
C, the unit purchase cost,
λ, the item's demand rate, and
T, the expected resupply time.

The parameters C, λ, and T are especially important since, as was discussed in section 2, the provisioning criterion is quite sensitive to them. This section is devoted to a discussion of possible ways to develop an item demand rate factor, λ, from the type of data currently available. Although the presentation is limited to the λ factor the techniques have application to the other factors.

5.1. Procedures for Estimating Demand Rates

Figure 6 illustrates the type of data that is available at the time of provisioning to estimate demand rates. Histograms of this type have been accumulated under a program started by Fleet Material Supply Office (FMSO) in 1963. The equipment groupings have been defined, for example, by placing all established stock list items into families by noun name and further grouping them into subfamilies according to application as denoted by the first two digits in the Component Identification Number (CID). The counts shown in the histogram represent information from equipment in that category. Individual counts can represent either Experienced Demand Replacement Factor (EDRF) or Technical Replacement Factor (TRF), e.g., engineering estimates or a combination of the two. Which type is actually chosen depends on the time in the field or age of the equipment and the equipment population size (for details see [1]). In Figure 6 the shaded portion represents demand rate estimates which have a TRF component; the other portions are pure EDRF estimates.

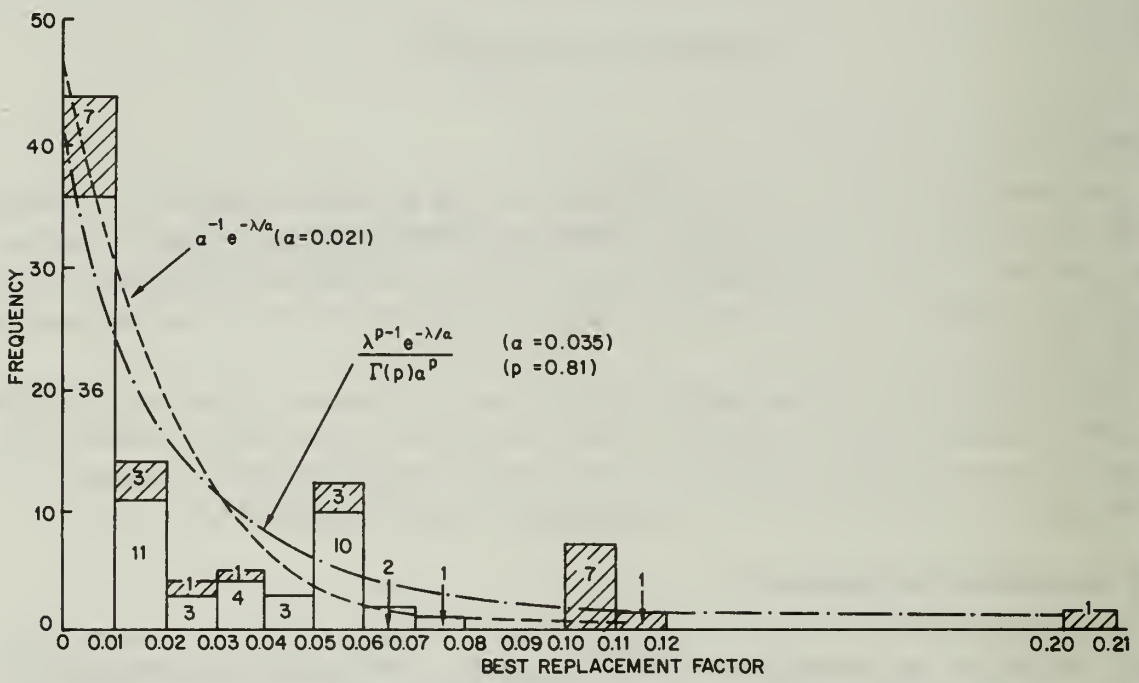


FIGURE 6. Best replacement factors for casing 01, 94 FSN's, MFRF = 0.028 (EDRF plain bar, TRF cross hatched bar) (from ALRAND 60[1], p. D-2).

The discussion which follows illustrates several methods for using this data to compute an item's demand factor. For this purpose the following generalized form of Equation (1) which includes the Bayes demand rate estimate will be used:

$$\gamma=[(1-\beta^{N})E_{\lambda}(1-e^{-\lambda T})]/[(1-\beta)C]. \tag{7}$$

To compute the complete provisioning criterion function, γ, and to test its sensitivity would require having the same type of histogram information on the cost, C, the expected resupply time, T, and the usable life for the FSN, N.

Several approaches are available for calculating the expected value $E_{\lambda}(e^{-\lambda T})$ from data of the type

indicated in Figure 6. The four approaches illustrated here are:

(i) by using the empirical histogram as a distributional estimate,
(ii) by using that gamma distribution estimate which best fits the histogram,
(iii) by using the mean of the data as a point estimate, and
(iv) by using the median of the data as a point estimate.

Each of these methods assumes something slightly different about the data. For example, approach (i) assigns equal weight to all available data points and uses a Bayes method to generate the best information on the demand rate. Probabilities derived from this histogram can easily be used to calculate the required value, i.e.,

$$E_\lambda(e^{-\lambda T}) = \sum_{i=1}^{n} e^{-\lambda_i T} p_i,$$

where

$p_i = \dfrac{\text{Number of observations in the } i\text{th interval}}{\text{Total number of observations}}$ and

λ_i = midpoint of the ith interval of the λ scale.

By using a gamma distribution, also a Bayes approach, one assumes the available data represent a random sample from a population of a particular form. The gamma distribution has the usual required properties for this purpose such as a continuous probability density function and allowing only positive values. This distribution which has two fitting parameters, also provides good approximations to many continuous distributions.

The point estimation methods assume

$$\hat{E}_\lambda(e^{-\lambda T}) \doteq e^{-\hat{\lambda} T},$$

where $\hat{\lambda}$ represents the particular point estimate of λ. Since this is not an equation but an approximation, the point estimation procedures are less desirable from a theoretical basis than the Bayes approach. For small demand rates, however, this approximation is almost an identity when λ is the mean.

Another estimate for λ is the median. Using the median value for $\hat{\lambda}$ has the effect of minimizing the impact of long tails (large demand rates). Since technical estimates in general tend to be too large, using the median has the benefit of minimizing their impact. However, because the number of technical estimates vary from item to item, the exact impact of this procedure on a general provisioning policy could be very questionable.

A comparison of these four approaches is illustrated with the data in Figure 6. Two examples are drawn from this data. One uses pure EDRF data only, the second uses all the data.

EXAMPLE 1 (*EDRF Data*):

The unshaded portion of Figure 6 shows the EDRF data. These data summarize the knowledge from actual experience about the Casing 01 type equipment demand rate. Table 2 presents the estimated

TABLE 2. *Summary of Best Replacement Factor Data*

	EDRF only		TRF and EDRF	
Interval	Number of observations	Prob. P_i	Number of observations	Prob. P_i
0.00–0.01	36	0.51	43	0.46
0.01–0.02	11	0.16	14	0.15
0.02–0.03	3	0.04	4	0.04
0.03–0.04	4	0.06	5	0.05
0.04–0.05	3	0.04	3	0.03
0.05–0.06	10	0.14	13	0.14
0.06–0.07	2	0.03	2	0.02
0.07–0.08	1	0.01	1	0.01
0.08–0.09				
0.09–0.10				
0.10–0.11			7	0.07
0.11–0.12			1	0.01
0.20–0.21			1	0.01
Number of observations	70		94	
Mean	0.021		0.028	
Median	0.010		0.013	
Standard deviation	0.021		0.031	

probability along with the summary statistics for the demand rate distribution for each BRF interval. Mean and standard deviations were computed by assuming probabilities to be massed at the center of the intervals.

The gamma distributions have the following probability density function:

$$g(\lambda \mid \alpha, p) = \frac{e^{-\lambda/\alpha}\lambda^{p-1}}{\alpha^p \Gamma(p)},$$

and the following characteristics:

(8) mean: $E(\lambda) = p\alpha$

(9) variance: $\mathrm{Var}(\lambda) = p\alpha^2$

expression in provisioning criterion: $E_\lambda(e^{-\lambda T}) = \left[\frac{1}{1+\alpha T}\right]^p.$

To estimate the gamma distribution for the current problem, note that estimates of the mean and standard deviation are equal. Application of Equations (8) and (9) yield $p = 1$ and $\alpha = 0.021$. In the case ($p = 1$) the Γ-density is the exponential curve shown on Figure 6 as the dashed curved.

The results of applying the Chi Square test to the exponential function shown in Figure 6 is presented in Table 3. The χ^2_{n-1} value computed, 17.2, is greater than the 99.9 percentage point of the Chi Square distribution. This would suggest the exponential curve is not a good fit of the observed data. However, the number of observations upon which this is based is not sufficient to give great confidence to this statement.

The impact on the provisioning factor, γ, (7) of using various techniques to incorporate demand rate is presented in Table 4. Recall that the demand rate, λ, influences this equation through the quantity $[1-E_\lambda(e^{-\lambda T})]$. The table shows, for various values of expected resupply time, T this quantity for point estimates (first two rows) and for demand rate distributions (last two rows). The numbers in parentheses represent the percent increase or decrease of the various estimates to the estimate based on the exponential distribution.

EXAMPLE 2 (*TRF and EDRF Data*):

This example uses all the data in Fig. 6 (EDRF data plus TRF). Summary statistics on this data are also presented in Table 3.

TABLE 3. *Chi Square Goodness of Fit Test*

Interval	Observed (O)	Expected (E)	$\frac{(O-E)^2}{E}$
0.00–0.01	36	26.5	3.4
0.01–0.02	11	16.5	1.8
0.02–0.05	10	20.5	5.5
0.05–	13	6.5	6.5
Total			17.2

χ^2_3 99.9 = 16.27

TABLE 4. $1-\mathrm{E}(\mathrm{e}^{-\lambda T})$ *Values for EDFR Data Only*

(Figures in parentheses indicate percent deviation from exponential result)

		Expected resupply time (T)		
		1 month	3 months	6 months
Point estimates	mean	0.00175 (0%)	0.00524 (0%)	0.01044 (0%)
	median	0.00083 (−52%)	0.0025 (−52%)	0.0050 (−52%)
Distribution estimates	Histogram	0.00167 (−5%)	0.00514 (−2%)	0.01027 (−1%)
	Exponential	0.00175 –	0.00522 –	0.01040 –

The shape of the histogram gives the general appearance of an exponential distribution; however, since the estimates of the mean and standard deviations are not equal, one might expect the two parameter gamma to provide a better fit.

Two procedures can be used to estimate the gamma parameters. The quick, although less accurate, way involves using the estimated mean and variance, see Equations (8) and (9). By using these two moments, it is seen that the gamma parameters satisfy

$$\alpha = \frac{\text{Var }(\lambda)}{E(\lambda)} \quad \text{and} \quad p = \frac{[E(\lambda)]^2}{\text{Var }(\lambda)}. \tag{10}$$

A more accurate procedure would be to use the maximum likelihood estimates (MLE). Details of how to derive MLE for the gamma parameters are given in Ref. [3]. The MLE distribution for the current problem is shown as the dotted curve in Figure 6.

Estimated values for $[1-E_\lambda(e^{-\lambda T})]$ by using the histogram, exponential, gamma (moments), and gamma (MLE) techniques are presented in Table 5. Values obtained by use of point estimates are also shown. Except when the median point estimate is used, the resulting value is rather insensitive to the procedure. Computationally the easiest method to apply when assuming a demand rate distribution involves the use of moment estimators (10) for a gamma distribution. Under this scheme, the estimate for $[1-E(e^{-\lambda T})]$ can be computed by using the following relationship

$$E(e^{-\lambda T}) = \left(\frac{1}{1+\dfrac{\text{Var }(\lambda)}{E(\lambda)}T}\right)^{\frac{[E(\lambda)]^2}{\text{Var }(\lambda)}}.$$

TABLE 5. $1-E(e^{-\lambda T})$ *Values for all Data*

(Figures in parentheses indicate percent deviation from Γ moment estimated results)

		Expected resupply time (T)		
		1 month	3 months	6 months
Point estimates	mean	0.0023 (0%)	0.0070 (0%)	0.0139 (0%)
	median	0.0011 (−52%)	0.00325 (−54%)	0.0065 (−53%)
Distribution estimates	Histogram	0.0024 (0%)	0.007389 (6%)	0.014716 (6%)
	Exponential	0.0023 (0%)	0.0070 (0%)	0.0138 (−1%)
	Γ distribution (moment est.)	0.0023 —	0.0070 —	0.0139 —
	Γ distribution (M.L.E.)	0.002 (0%)	0.0070 (0%)	0.0138 (−1%)

5.2. Findings and Conclusions

A good argument can be made for the use of the gamma distribution to factor demand rate into the provisioning decision. The estimated gamma parameters for this purpose should be based on current EDRF data from similar type equipment. The main advantage of a continuous distribution is that it smooths the raw data. Erroneous observations thus have less impact on the answers. Selection of the gamma distribution has several advantages. It has great flexibility in representing many kind of shapes, its range of possible values is the positive axis, and it yields simple formula for deriving $E_\lambda(e^{-\lambda T})$.

The reason for relying on only the EDRF data is the general tendency of demand rates based on TRF's to be unrepresentative of actual experience. Even worse, these estimates tend to be pessimistic, i.e., to overstate the demand rate. Thus their inclusion can have a sizable impact on the provisioning value γ as can be seen in the example of the previous section. In this example the estimates $[1-E(e^{-\lambda T})]$, and thus of the provisioning value λ, were between 30 and 35 percent greater based on all the data than when only EDRF data were used. Using only EDRF data may still overestimate the true family demand rate. This is caused by the tendency for the rather high TRF's to remain in the file longer and therefore to mask a larger percentage of low EDRF's.

To estimate the gamma parameters, it is suggested that the simple moment procedure (10) be used. Little, if any, increase in accuracy is gained through use of the more sophisticated maximum likelihood method. It is also worth noting in the example of the last section, that using the mean point estimate for demand rate gave essentially the same result as the Bayes procedures. Further sensitivity tests are indicated to determine if this estimate could be used in place of the gamma distribution without materially affecting the value of the provisioning criterion function.

APPENDIX

Transient Analysis

The expected level of backorders is not constant during a transient period which starts at the beginning of the item's life. Before ignoring this portion of an item's life in developing provisioning guidelines it is important to consider the behavior of backorders during this transient period, the probable length of this period, and the effect of ignoring the probability transient in developing a provisioning guideline. The results derived in this appendix concern the backorder levels expected during the transient period, the length of the transient period, and the error introduced by neglecting the transient period in analysis of backorders. The overall conclusion is that the transient period is of little consequence if the resupply time is small in comparison with the total useful life of the item. Thus the steady-state analysis is sufficient for developing insurance item provisioning guidelines.

MATHEMATICAL DEVELOPMENT

The development of the probabilistic behavior of backorders during the transient period is based on the mathematical properties of a filtered Poisson process. An excellent discussion of this process can be found in Parzen [4], chapter 4.

The general form for a filtered Poisson process, $X(t)$, is

$$X(t)=\sum_{n-1}^{N(t)} w(t, z_n, T_n),$$

where $N(t)$ is a Poisson process with parameter ν, z_n are the occurence times of the Poisson process and T_n are independent identically distributed random variables.

The probabilistic properties of $X(t)$ are

$$\text{characteristic function: } \Phi_{X(t)}(u) = \exp\left[\nu \int_0^t E\left[e^{iuw(t,z,T)} - 1\right] dz\right],$$

$$\text{mean: } E(X(t)) = \nu \int E[w(t, z, T)]\, dz,$$

and

$$\text{variance: Var } (X(t)) = \nu \int_0^t E[w^2(t, z, T)]\, dz.$$

In the current application,

$$w(t, z_n, T_n) = w(t - z_n, T_n) = \begin{cases} 1 \text{ if } T_n > t - z_n \\ 0 \text{ otherwise,} \end{cases} \tag{11}$$

where z_n's represents the times demands occur, and T_n represents the resupply time for the nth item. $X(t)$ thus represents the number of orders outstanding against the equipment at time t. The number of backorders given a stocking policy level of s is

$$B(t) = \begin{cases} X(t) - s \text{ if } X(t) > S \\ 0 \text{ otherwise.} \end{cases}$$

For this form of $w(t, z, T)$ the Equations in (11) reduce to

$$\Phi_{X(t)}(u) = \exp\left[\nu\,(e^{iu} - 1) \int_0^t P(T > z)\, dz\right]$$

$$E(X(t)) = \nu \int_0^t P(T > z)\, dz$$

(12)

$$\text{Var } (X(t)) = \nu \int_0^t P(T > z)\, dz.$$

The formula for $\Phi_{X(t)}(u)$ implies that $X(t)$ at any given t has a Poisson distribution with parameter $\nu \int_0^t P(T > z)\, dz$. (Note as $t \to \infty$, this becomes $\nu E(T)$, the steady-state result.)

DURATION OF TRANSIENT STATE

How long does the transient state last? As seen by the equations in (12) the answer to this question depends upon the resupply time distribution. To investigate typical behavior, we shall consider several possible resupply time distributions.

For convenience we shall let $R(z) = P(T > z)$, and $L(t) = \int_0^t R(z)\, dz$.

EXAMPLE: *Uniform distribution* (T ~ U [0, 2E(T)])

$$\text{Var } (T) = \frac{(ET)^2}{3}$$

$$L(t) = \int_0^t R(z)\, dz = \int_0^t \frac{2E(T) - z}{2E(T)}\, dz$$

$$= \frac{1}{2\ ET}\left(2t\ ET - \frac{t^2}{2}\right) \qquad \text{for } 0 \leqslant t \leqslant 2\ ET.$$

A comparison of t and $L(t)$ are given in Table 6.

TABLE 6. *Comparison of Transient State to Steady State at Various Points of Time (Uniform Distribution)*

(t)	$L(t)/E[T]$
$ET - \sqrt{\text{Var } T}$	0.372
ET	0.750
$ET + \sqrt{\text{Var } T}$	0.955

EXAMPLE: *Gamma distribution* (T ~ $\Gamma(\alpha,\text{p})$)

$$ET = p/\alpha$$

$$\text{Var } T = p/x^2.$$

Again,

$$L(t) = \int_0^t R(x)\, dx = t(1 - \Gamma(t)) + \int_0^t x d\Gamma(x).$$

Values for the cumulative Γ distribution function can readily be determined by using tables of the cumulative Chi Square distribution, X^2. The relationship between the two distributions is

$$\Gamma_{\alpha,p}(t) = \chi^2_{2p}\left(\frac{2p}{ET}\, t\right).$$

Thus

$$L(t) = t\left[1 - \chi^2_{2p}\left(\frac{2pt}{ET}\right)\right] + ET\ \chi^2_{(p+1)}\left(\frac{2pt}{ET}\right).$$

Note the above expression for $L(t)$ depends only on the gamma parameter, p, when the expected re-

supply time, ET, is fixed. The value p is called a shape parameter by some authors and relates to the amount of dispersion in the distribution, i.e.,

$$\text{Coefficient of variation} = \sqrt{\frac{\text{Var }(\cdot)}{\text{E }(\cdot)}} = \frac{1}{\sqrt{p}} .$$

Thus the larger the p value the less disperse the distribution.

$L(t)$ values for the gamma distribution and various coefficients of variation are given in Table 7.

TABLE 7. *Comparison of Transient State to Steady State at Various Points in Time* (Γ *Distribution*).

t	$L(t)/E(T)$		
	Coef. of Variation		
	0.5 (p=4)	1 (p=1)	1.414 (p=1/2)
1/2 ET	0.47	0.4	0.32
ET	0.81	0.63	0.5
$2ET$	0.98	0.88	0.74
$3ET$	0.999	0.95	0.87

Finally note that as the coefficient of variation goes to zero

$$L(t) \to \begin{cases} t \text{ if } u \leq E(T) \\ ET \text{ if } u > ET. \end{cases}$$

This follows since $L(t) = \int_0^t P(T > z)\, dz$, and

$$P(T > z) \to \begin{cases} 1 \text{ if } z \leq ET \\ 0 \text{ if } z > ET \end{cases} \quad \text{as the coefficient of variation} \to 0.$$

IMPLICATION FOR PROVISIONING

Assume that the time between replacement demands has an exponential probability distribution with demand rate λ; then the number of demands, $N(t)$, that occur up to a given time t has a Poisson distribution with parameter λt. Under this assumption the number of backorders that exist against the system at any point in time, t, follows a Poisson distribution with parameter:

$$\lambda \int_0^t P(T > z)\, dz ,$$

where T is the random variable denoting the resupply time. As t increases, $\int_0^t P(T > z)\, dz \to E(T)$.

When $\int_0^t P(T > z)\, dz \approx E(T)$ we say that the backorder behavior is in a steady state.

The duration of the transient state is evidently related to the resupply time distribution. This relationship was investigated above for several specific resupply time distributions. The results of this investigation are presented in Figure 7, and lead to the following conclusions:

- The length of the transient period is closely related to the mean (average) resupply time.
- Given the mean resupply time the length of the transient period is mainly affected by the coefficient of variation, $\frac{\text{standard deviation}}{\text{mean}}$, of the resupply time. The length of the transient increases as the coefficient increases.
- In the case of constant resupply time (zero coefficient of variation) the transient period terminates after the lapse of one resupply time.
- For coefficients of variation less than 0.5, the process reaches steady state after about 1.5 mean resupply times. In the case of larger coefficients longer times are required to achieve steady state. This time, however, is still measured in terms of only a few mean reorder times.

These results concerning the length of the transient period have a strong impact on the development of initial provisioning procedures. The conclusion that can be reached is that the difference between an exact analysis of backorders and an approximation using only steady-state results is small. For example, given a typical resupply time for an item of say 0 to 180 days and life expectancy of 8 years, the transient period exists for only about 5 percent of the total life. This implies that steady-state estimated *backorder-years* is only about 3 percent more than the true value and that the transient can therefore be ignored in developing an initial provisioning stocking rule.

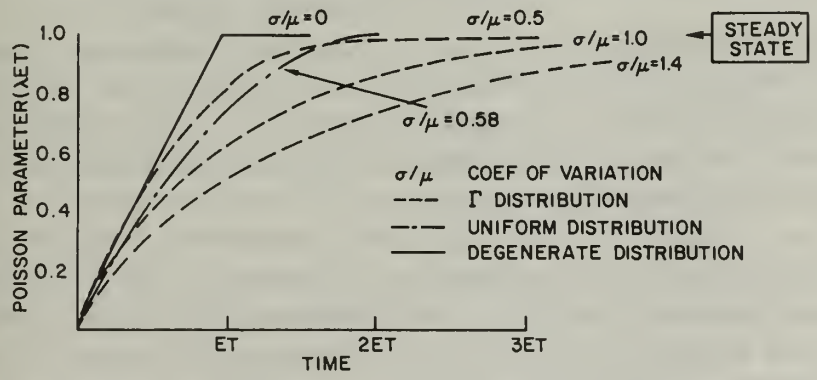


FIGURE 7. Transient nature of backorders.

REFERENCES

[1] ALRAND Report 60, Replacement Factors for Repair Parts, Operations Analysis Department, U.S. Navy Fleet Material Support Office, Mechanicsburg, Pennsylvania (19 July 1968).

[2] Burton, R. W. and S. C. Jaquette, The Initial Provisioning Decision For Insurance Type Items. Nav. Res. Log. Quart. *20*, 123–146 (1973).

[3] Choi, S. C. and R. Wette, Maximum Likelihood Estimation Of The Parameters Of The Gamma Distribution and Their Bias, Technometrics *11*, 683–690 (1969).

[4] Parzen, E., *Stochastic Processes* (Holden-Day, San Francisco, 1962).

PERIODIC REPLACEMENT WITH MINIMAL REPAIR AT FAILURE AND ADJUSTMENT COSTS*

C. Tilquin and R. Cléroux
Université de Montréal
Département d'Informatique

ABSTRACT

We investigate periodic replacement policies with minimal repair at failure, thereby, minimizing the average expected cost per unit time over an infinite time span. The standard cost structure is modified by the introduction of a term which takes adjustment costs into account.

1. INTRODUCTION

This paper is concerned with periodic replacement policies with minimal repair at failure. In those policies a unit is replaced at age T. If it happens to fail between the periodic replacements, it is repaired in such a way that the unit failure rate remains undisturbed. This replacement strategy has been discussed in [3] and [4] with a cost structure which considers only a cost c_1 for repairing a failed unit and a cost c_2 of periodic replacement at age T.

More recently, more sophisticated cost structures have been considered in [1] and [2] for the age replacement problem. The motivation behind the introduction of such cost structures is to be able to cope with more realistic situations. The main advantage of the classical structure is its simplicity which permits easy mathematical formulation, analysis, and even comparison between various preventive models. From this point of view, the standard approach is important because it gives us a strong theoretical basis for further developments. Practically, however, its simplicity is a weakness because the elasticity of the "c_1 and c_2" structure is very limited so that it is often unable to deal with real cases. That is the reason why, in considering here the preventive replacement model with minimal repair at failure, our purpose will be double. We want first to introduce in this model a cost structure closer to reality and secondly, we want the new model to remain simple enough to be easily manageable mathematically. It appeared to us that the cost structure proposed in [1], because of its general character, had the elasticity necessary to approach reality. It follows from the discussion presented in [1], that most of the costs incurred in real situations can be represented adequately by this discrete costs structure—that is, for example, periodic adjustment costs, depreciation costs, or interest charges. They will all be referenced in the following as "adjustment costs." It will also appear to the reader that this mathematical model corresponding to that cost structure remains sufficiently simple to give results of practical use.

* The research of this paper was supported in part by the Defence Research Board of Canada, Grant Number DRB9740–18(UG).

In this paper we consider the periodic replacement model with minimal repair at failure with the cost structure from [1]. One should be equally justified to use the cost factors from [2], but this will not be considered in this paper. Our problem here can be stated as follows:

A unit is replaced at times $T, 2T, 3T \ldots$. Between these periodic replacements it may happen to fail and it is immediately repaired in such a way that its failure rate $r(t)$ (corresponding to a time to failure distribution F having continuous density f, finite mean and $F(0)=0$) is not disturbed. $r(t)$ is assumed to be continuous and differentiable.

It costs c_1 to repair a failed unit and c_2 to replace it at age T. A cost $c_3(ik)$ is suffered at age ik, $i=1, 2, 3, \ldots; k>0$. All the costs are assumed to be greater than or equal to zero. We want to find a policy which minimizes the expected cost per unit time over an infinite time span.

2. NOTATION AND BASIC RESULTS

Let $N_1(t)$ denote the number of failures in $[0, t]$ and $N_2(t)$ denote the number of preventive replacements in $[0, t]$.

The preventive replacements are performed at fixed intervals of length T. Then

$$\lim_{t\to\infty}\frac{EN_2(t)}{t}=\frac{1}{T}, \tag{2.1}$$

where E denotes expectation.

It has been proved in [3] that

$$EN_1(T)=\int_0^T r(u)du,$$

where $r(t)$ is the unit failure rate at time t.

Let $N_{1i}(T)$ be the number of failures in the interval $[(i-1)T, iT)$. $N_{11}(T), N_{12}(T), \ldots$ are independent and identically distributed random variables. By the law of large numbers

$$\lim_{k\to\infty}\frac{\sum_{i=1}^{k} N_{1i}(T)}{k}=EN_1(T) \quad \text{a.s. (almost surely).} \tag{2.2}$$

If $kT\leq t<(k+1)T$ then

$$\frac{\sum_{i=1}^{k} N_{1i}(T)}{kT\frac{(k+1)}{k}}\leq\frac{N_1(t)}{t}<\frac{\sum_{i=1}^{k+1} N_{1i}(T)}{(k+1)T\frac{k}{(k+1)}}$$

By taking the limit and using (2.2), we get

$$\lim_{t\to\infty}\frac{N_1(t)}{t}=\frac{EN_1(T)}{T} \quad \text{a.s.} \tag{2.3}$$

In the same way $E\sum_{i=1}^{k} N_{1i}(T) = \sum_{i=1}^{k} EN_{1i}(T) = kEN_1(T)$ implies that

$$\lim_{t\to\infty} \frac{EN_1(t)}{t} = \frac{EN_1(T)}{T} = \frac{\int_0^T r(u)du}{T}. \tag{2.4}$$

Finally, let $C(t)$ denote the cost of operating the system up to time t and let $C(T) = \lim_{t\to\infty} \frac{EC(t)}{t}$ denote the long run average expected cost per unit time for a policy of period T. The problem will be to find the policy T which minimizes $C(T)$.

3. THE STANDARD COST STRUCTURE

The periodic replacement model with minimal repair at failure is discussed in [3] and [4] for the standard cost case. The problem is to select T so as to minimize

$$C(T) = \lim_{t\to\infty} \frac{EC(t)}{t} = \lim_{t\to\infty} \frac{c_1EN_1(t) + c_2EN_2(t)}{t}.$$

Using section 2 results, we get

$$C(T) = \frac{c_1\int_0^T r(u)du + c_2}{T}. \tag{3.1}$$

T minimizing $C(T)$ must be a solution of $\frac{\delta C(T)}{\delta T} = 0$. This can be written as

$$L(T) \equiv \int_0^T [r(T) - r(u)]du = \frac{c_2}{c_1}. \tag{3.2}$$

It is proved in [4] that:

If there exists $[0, b)$ where $0 < b \leq \infty$ such that $r(t)$ is continuous and unbounded on $[0, b)$, then (3.2) has a solution T_0^* in T as $L(T)$ assumes all nonnegative values when T ranges from 0 to b.

This solution T_0^* is unique if $r(t)$ is strictly increasing and differentiable.

4. DISCRETE ADJUSTMENT COSTS

We consider now a cost structure involving discrete costs of keeping the system in operation together with the standard costs c_1 and c_2.

Here, a known cost $c_3(ik)$ is suffered at age ik, $i = 1, 2, \ldots$ where $k > 0$, $c_3(0) = 0$ and $c_3(s) \geq 0$ for any $s = ik$. There is no other assumption on the sequence $c_3(0)$, $c_3(k)$, $c_3(2k)$, . . . and in particular we do not assume that this sequence is increasing. Let

$$v(T) = \left[\frac{T}{k}\right] - \delta_{\left[\frac{T}{k}\right]\cdot k, T}$$

where δ_{ij} is the Kronecker's delta and $[y]$ is the largest integer smaller than or equal to y, for any y. Let $C^*(v(T))$ be the total adjustment cost over a unit life of length T, that is

$$C^*(v(T)) = \sum_{i=0}^{v(T)} c_3(ik).$$

The definition of $v(T)$ and the assumption $c_3(0)=0$ imply that any adjustment cost suffered a the beginning or the end of a unit life is taken into account in the replacement cost c_2. $C^*(v(T))$ is step function jumping at points ik, $i=1, 2, 3, \ldots$ and $A(t)$, the total adjustment cost up to time t, is

$$A(t) = rC^*(v(T)) + C^*(v(\Delta t)),$$

if $rT \leq t < (r+1)T$ and $\Delta t = t - rT$.

Then the total cost of operating the system up to time t is:

$$C(t) = c_1 N_1(t) + c_2 N_2(t) + rC^*(v(T)) + C^*(v(\Delta t)).$$

By noting that $\lim_{t\to\infty} \frac{A(t)}{t} = \frac{C^*(v(T))}{T}$ and using section 3 results, we get

$$C(T) = \frac{c_1 \int_0^T r(u)\,du + c_2 + C^*(v(T))}{T}. \tag{4.1}$$

Let $C(T) = C_1(T) + C_2(T)$ with

$$\begin{cases} C_1(T) = \dfrac{c_1 \int_0^T r(u)\,du + c_2}{T} \\ C_2(T) = \dfrac{C^*(v(T))}{T}. \end{cases} \tag{4.2}$$

We have that $C_1(T)$ results only from the standard costs c_1 and c_2 and is continuous since $r(-)$ is continuous by assumption; we then get

(i) $C_1(0^+) = \infty$.

(ii) $C_1(\infty) = c_1 r(\infty)$ (by using l'Hospital's rule), when $\int_0^\infty r(u)\,du = \infty$, which is always the case in practice.

(iii) $C_1'(T) = \frac{c_1 L(T) - c_2}{T^2}$ which implies that

$$\left\{ C_1'(T) \gtrless 0 \right\} \text{iff} \left\{ L(T) \gtrless \frac{c_2}{c_1} \right\}.$$

(iv) $C''(T) = \{c_1[L'(T) \cdot T - 2L(T)] + 2c_2\}/T^3$ if $r(T)$ is differentiable, which implies that

$$\{r(T) \text{ (strictly) convex}\} \text{ iff } \{C_1(T) \text{ (strictly) convex}\}.$$

We also have that $C_2(T)$ is a piecewise continuous function of T with discontinuities at points ik, $i=1, 2, \ldots$ but is continuous from the left at these points with $C_2(ik) \leq C_2(ik^+)$. $C_2(T)$ is differentiable on each of its continuity intervals $ik < T < (i+1)k$, $i=0, 1, \ldots$ where $C^*(v(T))$ is a constant. Then

$$C_2'(T) = -\frac{C^*(v(T))}{T^2} < 0$$

and

$$C_2''(T) = \frac{2C^*(v(T))}{T^3} > 0,$$

except on the first interval $0 < T < k$ where $C_2'(T) = C_2''(T) = 0$. From this we can say that except on the first interval where it is identically null, $C_2(T)$ is a strictly convex decreasing function on each of its continuity intervals. We also have that

$$C_2(0^+) = 0 \qquad \text{and} \qquad C_2(\infty) = \lim_{x \to \infty} \frac{C^*(v(x))}{x} = \gamma,$$

say, and if $\beta = \lim_{x \to \infty} C^*(v(x)) < \infty$, then $\gamma = 0$ and $C_2(\infty) = 0$.

Combining these results for $C(T) = C_1(T) + C_2(T)$ yields statements (4.3) and (4.4) below

(4.3)
$$\begin{cases} C(0^+) = \infty \\ C(\infty) = c_1 r(\infty) + \gamma \qquad \text{if } \int_0^\infty r(u)\,du = \infty \end{cases}$$

(Note that $r(\infty) = \infty$ implies that $C(\infty) = \infty$.)

(4.4)
$$\begin{cases} C(T) \text{ is piecewise continuous with jumps at } ik,\ i=1, 2, \ldots \text{ where it is left continuous} \\ \quad \text{with } C(ik) \leq C(ik^+). \text{ On each interval } ik < T < (i+1)k,\ i=1, 2, \ldots, \\ C(T) \text{ is strictly convex iff } r(t) \text{ is convex.} \\ \text{On } 0 < T < k,\ C(T) \text{ is (strictly) convex iff } r(t) \text{ is (strictly) convex.} \end{cases}$$

A necessary condition for the existence of a local minimum of $C(T)$ for $T \neq vk$, v positive integer, is that $C'(T)=0$; by (4.4), this condition is also sufficient if $r(t)$ is convex. If a local minimum of $C(T)$ is obtained at $T^* \neq vk$, $v=1, 2, \ldots$, then T^* has to satisfy

$$L(T)=\frac{c_2+C^*(v(T))}{c_1}. \tag{4.5}$$

Then

$$C^*(v(T^*))=c_1L(T^*)-c_2,$$

and by replacing $C^*(v(T^*))$ with its value in (4.1) gives

$$C(T^*)=c_1r(T^*). \tag{4.6}$$

If $r(t)$ is convex, then it follows from (4.4) that for $T>k$, a local minimization point of $C(T)$ on a continuity interval is unique and thus global on this interval; for $0<T<k$ any local minimization point is also global, but not necessarily unique on this interval except when $r(t)$ is strictly convex.

We are now in a position to prove the following theorems.

THEOREM (4.1): The absolute minimum of $C(T)$ exists on $[0, \infty]$ and is reached at one of the points ik, $i=1, 2, 3, \ldots$ or at one of the local minimization points T_i^* of $C(T)$, $i=1, 2, \ldots$.

PROOF: This comes from the fact that $C(T)$ is piecewise continuous, left continuous at ik, $i=1, 2, 3, \ldots$ with $C(ik^+)>C(ik)$.

THEOREM (4.2): If $r(t)$ is increasing, the minimum of $C(T)$ is reached at one of the points ik, $i=1, 2, \ldots$ and $T_1^* \neq ik$, the first of the local minimization points of $C(T)$.

PROOF: If $T_1^* \neq ik$, Theorem (4.2) results from Theorem (4.1) and from Equation (4.6) by considering the fact that $r(t)$ is assumed increasing.

THEOREM (4.3): The existence of an interval $[0, T_0)$, $0<T_0<\infty$ such that, for $t \in [0, T_0)$, $r(t)$ is unbounded, $r'(t)<\infty$ and $\lim_{t \to T_0} C^*(v(t))<\infty$, is a sufficient condition for the existence of a solution T^* of (4.5) with $T^* \in [0, T_0)$, $T^* \neq ik$, $i=1, 2, 3, \ldots$. Moreover, the existence of T_0^*, the solution of (3.2) in the standard case, is a necessary condition for the existence of T^*.

PROOF: The first part of the theorem can be proved in a way similar to that followed in [3] for proving Theorem (2.5). It is based on the fact that $L(T)$ assumes all nonnegative values when T goes from 0 to T_0. At the same time the right-hand member of (4.5) is an increasing step-function which is bounded, by the assumption that $\lim_{T \to T_0} C^*(v(T))<\infty$. The second part of the theorem results from the first part as Equation (4.5) corresponds to the standard case when $C^*(v(T))=0$. Then obviously T_0^* exists when T^* exists and $T^* \geq T_0^*$.

COROLLARY (4.1): If $r(t)$ is convex, the existence of an interval $[0, T_0)$, $0<T_0 \leq \infty$, such that for $t \in [0, T_0)$, $r(t)$ is unbounded, $r'(t)<\infty$ and $\lim_{t \to T_0} C^*(v(t))<\infty$ is a sufficient condition for the existence of a local minimization point T^* of $C(T)$ with $T^* \in [0, T_0)$, $T^* \neq vk$, $v=1, 2, \ldots$.

Moreover, the existence of an optimal solution T_0^* in the standard cost case is a necessary condition for the existence of T^*.

PROOF: That results directly from Theorem (4.3), from (4.4) and, for the last part, from the fact hat $C_1(T)$ is convex when $r(T)$ is convex.

THEOREM (4.4): Two sufficient conditions for the existence of a finite minimization point of $C(T)$ are $\gamma = \infty$ and $r(\infty) = \infty$. Moreover, for $\gamma \neq \infty$ and $r(\infty) \neq \infty$, a sufficient condition is that, when $T \to \infty$, $C(T)$ converges from below to

$$C(\infty) = \begin{cases} c_1 r(\infty) & \text{if } \int_0^\infty r(u)du = \infty \quad \text{and } \gamma = 0 \\ \gamma + c_1 r(\infty) & \text{if } \int_0^\infty r(u)du = \infty \quad \text{and } \gamma \neq 0. \end{cases}$$

PROOF: The first part of the theorem follows from (4.3) and Theorem (4.1) since we know that $C(0) = \infty$ and that $\gamma = \infty$ or $r(\infty) = \infty$ imply that $C(\infty) = \infty$. When $\gamma \neq \infty$ and $r(\infty) \neq \infty$, then $C(\infty) \neq \infty$. And since $C(0) = \infty$, the second part of the theorem also follows from (4.3) and Theorem (4.1).

5. THE SEARCH FOR AN OPTIMAL SOLUTION

In this section, we give some theoretical results which are of some importance from a practical point of view. They can be very useful in the search for an optimal solution because they define or characterize a limited set of possible solutions in which we are ensured to find an optimal one. In fact, the following theorems bring us efficient stopping rules for our search procedure.

THEOREM (5.1): If there exists a T_0 such that, for all $T \geq T_0$,

$$C^*(v(T)) < c_1 L(T) - c_2, \tag{5.1}$$

hen $C(T)$ has a minimization point $\leq T_0$.

PROOF: We have

$$C'(T) = \frac{c_1 L(T) - c_2 - C^*(v(T))}{T^2}.$$

f $C'(T) > 0$ for all $T \geq T_0$, $T \neq vk$, $v = 1, 2, 3, \ldots$, the minimum of $C(T)$ is reached at $T \leq T_0$ by Theorem (4.1) since $C(vk^{\pm}) \geq C(vk)$. But $C'(T) > 0$ is equivalent to

$$C^*(v(T)) < c_1 L(T) - c_2.$$

COROLLARY (5.1): If $\gamma = 0$, if $r(t)$ is increasing and if there exists T_0 such that $L(T_0) > \frac{\beta + c_2}{c_1}$, hen $C(T)$ reaches its minimum at some $T \leq T_0$.

PROOF: By differentiating $L(T)$, we get $L'(T) = Tr'(T)$ and so $L(T)$ is increasing when $r(T)$ is ncreasing. Then, if for some T_0, $L(T_0) > \frac{\beta + c_2}{c_1}$, we also have that $c_1 L(T_0) - c_2 > \beta$. Finally, the fact hat $\beta \geq C^*(v(T))$, for all T and that $L(T)$ is increasing imply that $c_1 L(T) - c_2 > C^*(v(T))$ for all $T \geq T_0$.

COROLLARY (5.2): If $\frac{C^*(v(T))}{T}$ converges to γ from below as T goes to ∞ and if $\lim_{T\to\infty} \frac{L(T)}{T} \geq \frac{\gamma}{c}$ then $C(T)$ reaches its minimum at a finite point.

PROOF: We want that $C^*(v(T)) < c_1 L(T) - c_2$ for all $T \geq T_0$ for T_0 finite. This can also be written as

$$\frac{C^*(v(T))}{T} < \frac{c_1 L(T)}{T} - \frac{c_2}{T} \qquad \text{for all } T \geq T_0.$$

If $\frac{C^*(v(T))}{T}$ converges from below to γ, it is then sufficient to verify this inequality at the limit, that is

$$\frac{\gamma}{c_1} < \lim_{T\to\infty} \frac{L(T)}{T}.$$

Example; As an illustration of Theorem (5.1) and its corollary, consider the following artificial example:

Suppose $F(t)$ is Weibull with $\lambda > 0$ and $\alpha = 2$; then $r(t) = 2\lambda t$. Adjustment costs are supposed constant $c_3(ik) = c_3$, $i = 1, 2, \ldots$ and thus $C^*(v(T)) \leq \frac{c_3 T}{k}$. Here $L(T) = \lambda T^2$ and Theorem (5.1) states that the minimum of $L(T)$ is reached at $T \leq T_0$, if for all $T \geq T_0$

$$\frac{c_3 T}{k} < c_1 \lambda T^2 - c_2$$

or

$$k\lambda c_1 T^2 - c_3 T - c_2 k > 0.$$

This inequality is true for $T \notin [T_1', T_2']$, where T_1' and T_2' are the roots of the above second degree polynomial. Thus, it is sufficient to choose T_2' for T_0, and $C(T)$ has a minimization point $\leq T_0$ with

$$T_0 = \frac{c_3 + \sqrt{c_3^2 + 4c_2 c_1 k^2 \lambda}}{2k\lambda c_1}.$$

On the other hand, $\gamma = \lim_{T\to\infty} \frac{C^*(v(T))}{T} \leq \lim_{T\to\infty} \frac{c_3 T}{kT} = \frac{c_3}{k}$ and $\lim_{T\to\infty} \frac{L(T)}{T} = \lim_{T\to\infty} \lambda T > \frac{c_3}{c_1 k} \geq \frac{\gamma}{c_1}$. Corollary (5.2) condition is verified, which confirms the above result.

The rule obtain below is useful as a stopping rule in the search for a finite optimum.

THEOREM (5.2): If $r(t)$ is increasing, then the existence of $v_0 > 1$ such that

$$\int_0^{vk} r(u)\,du - v \int_{vk}^{(v+1)k} r(u)\,du \leq \frac{vc_3(vk) - C^*(v-1) - c_2}{c_1}$$

for all $v > v_0$, and with strict inequality for $v = v_0$, is a necessary and sufficient condition for the minimum of $C(T)$ to be reached at one of the points vk, $v = 1, 2, \ldots, v_0$ and T_1^* if T_1^* exists, or at one o

he points vk, $k = 1, 2, \ldots, v_0$ if not.

PROOF: Suppose there exists a $v_0 > 1$ with $C(vk) \leq C((v+1)k)$ for all $v > v_0$ with strict inequality when $v = v_0$. We know that $C(vk) \leq C(vk^+)$. Then the above assumption implies that $\min_T C(T)$ is obtained at a finite point. This assumption can also be written as

$$\frac{c_1 \int_0^{vk} r(u)du + c_2 + C^*(v-1)}{vk} \leq \frac{c_1 \int_0^{(v+1)k} r(u)du + c_2 + C^*(v)}{(v+1)k},$$

with a strict inequality for $v = v_0$. The last inequality is equivalent to

$$\frac{c_1 \int_0^{vk} r(u)du + c_2 + C^*(v-1)}{c_1 \int_{vk}^{(v+1)k} r(u)du + c_3(vk)} \leq v$$

or

$$\int_0^{vk} r(u)du - v \int_{vk}^{(v+1)k} r(u)du \leq \frac{vc_3(vk) - C^*(v-1) - c_2}{c_1}.$$

COROLLARY (5.3): If $r(t)$ is increasing, then $c_3(vk) \geq \frac{c_2 + C^*(v-1)}{v}$ for all $v > v_0 > 1$ with strict inequality for $v = v_0$ is a sufficient condition for $\min_T C(T)$ be obtained at vk, $v = 1, \ldots, v_0$ and T_1^* if it exists.

COROLLARY (5.4): If F is exponential, then $c_3(vk) \geq \frac{c_2 + C^*(v-1)}{v}$ for all $v > v_0 > 1$ with strict inequality for $v = v_0$ is a necessary and sufficient condition for $\min C(T)$ be obtained at vk, $v = 1, \ldots, v_0$.

COROLLARY (5.5): If F is *IFR*, if T_0^*, the optimal solution in the standard model, exists and is such that $(v_0 - 1)k < T_0^* \leq v_0 k$, if there exists a $v_1 > 1$ such that $C^*(v-1) \leq vc_3(vk)$ for all $v > v_1$ and strict inequality for $v = v_1$ then $\min_T C(T)$ is obtained at one of the points T_1^* (if it exists), vk, $v = 1, 2, \ldots \max(v_0, v_1)$.

PROOF: Recall that $C(T) = C_1(T) + C_2(T)$. Then T_0^*, it if exists, is a solution of $L(T) = \frac{c_2}{c_1}$. We have seen that $L(T)$ is increasing if $r(t)$ is increasing and that $C_1'(T) > 0$ is equivalent to $L(T) > \frac{c_2}{c_1}$. Then $C_1(T)$ is increasing when $T \geq T_0^*$. Now if we want that $C_2(vk) \leq C_2((v+1)k)$ for all $v > v_1$ and strict inequality for $v = v_1$ we must verify that $(v+1)C^*(v-1) \leq vC^*(v)$ or $C^*(v-1) \leq vc_3(vk)$ for all $v > v_1$ and strict inequality for v_1. Therefore $C(vk) \leq C((v+1)k)$ for all $v > \max(v_0, v_1)$ with strict inequality for $\max(v_0, v_1)$.

6. EXAMPLES

We now illustrate the search for a solution and the behavior of both the standard cost $C_S(T)$ and our modified cost $C(T)$. We have used examples taken from [1] and computed optimal costs and

policies using a program written in FORTRAN for the CDC CYBER 74 computer Université de Montréal. Our results are compared with the results in [1] for the age replacement policy.

Consider the following examples taken from [1]. The unit life is Weibull with parameters $\alpha = 3.3$ and $\beta = 1$. We assume that depreciation costs are incurred each k units of time and we will consider $k = 0.2, 0.02, 0.05$. The depreciation costs are of the following form:

$$c_3(k) = p \qquad \text{and} \qquad c_3(ik) = s \cdot c_3((i-1)k) \qquad \text{for } i = 2, 3, \ldots .$$

In our examples $s = 0.98$ and 1.40, $p = 50.0$ and 200.0. Let T^* be the optimal replacement policy and $C(T^*)$ the associated minimum cost. The following table summarizes our results:

Example	c_1	c_2	p	k	s	$C(T^*)$	T^*
1	320	30	50	0.2	0.98	157.89	0.20
2				0.02	0.98	1,500.09	0.02
3				0.05	1.40	600.32	0.05
4	1,100	100	50	0.2	0.98	508.69	0.40
5				0.02	0.98	2,343.18	0.52
6				0.05	1.40	1,480.25	0.15

Examples 1 and 4 are illustrated in Figures 1 and 2 together with the standard cost function $C_S(T)$ which is minimum in T_0^*. On the other hand, T_1^* is the first local minimum of $C(T)$ satisfying

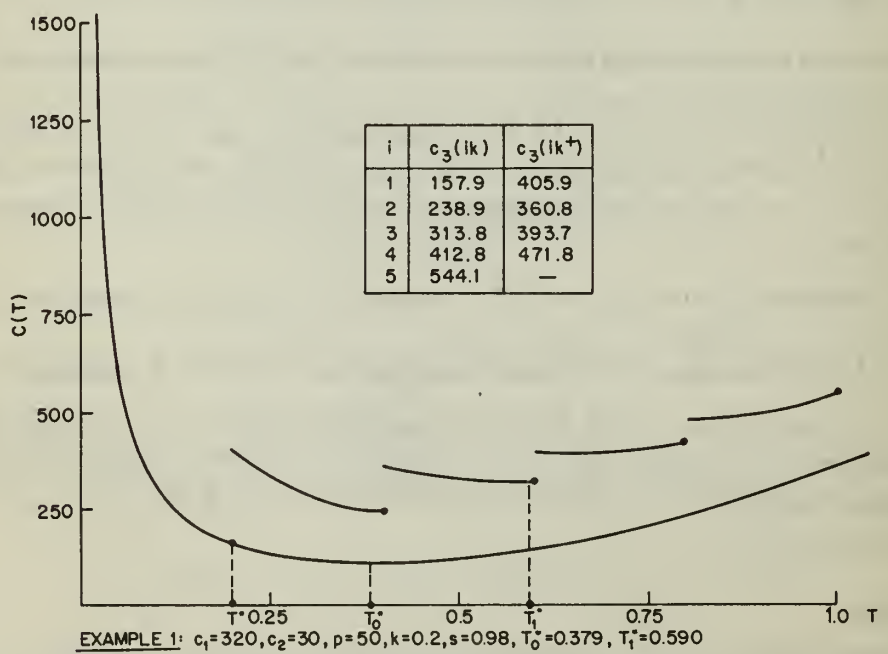
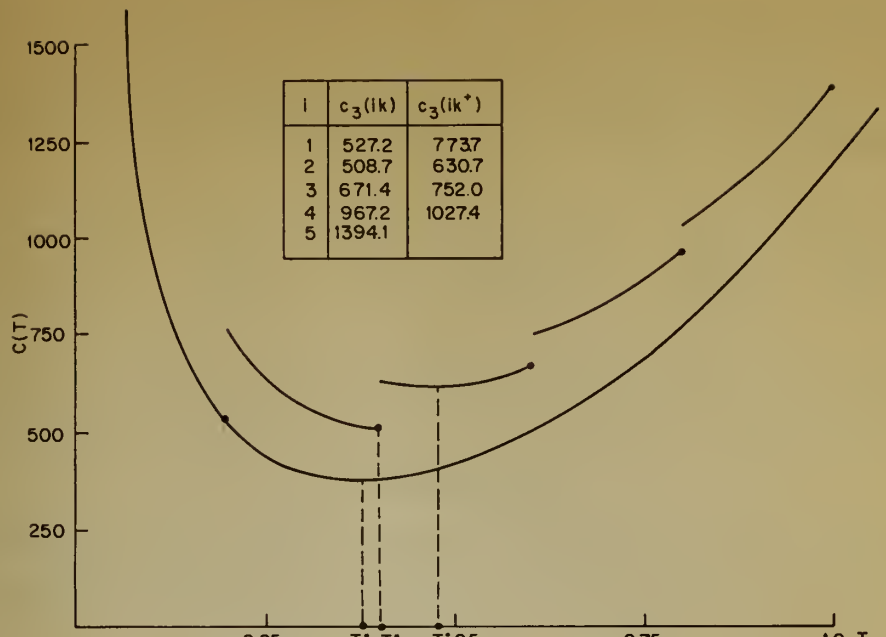


FIGURE 1

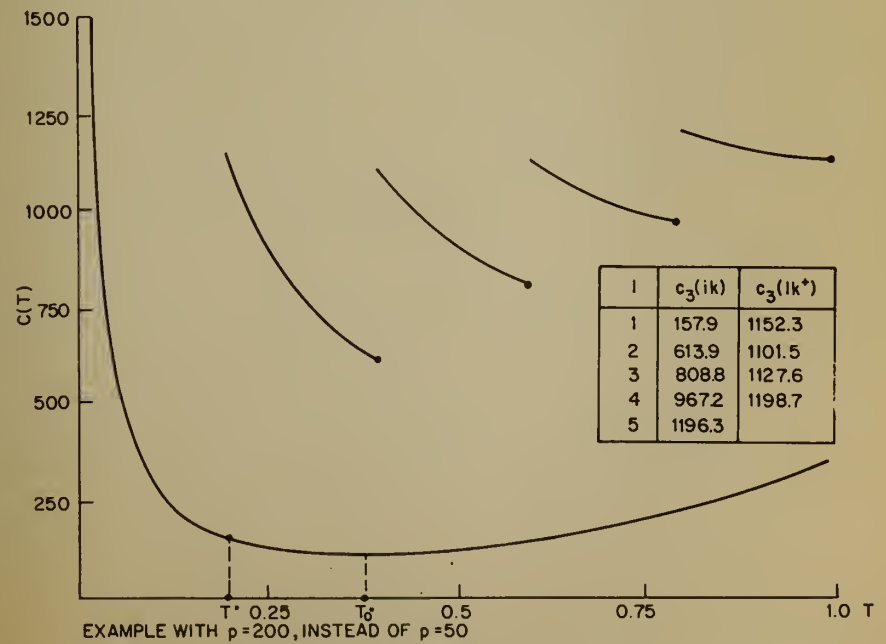


i	$c_3(ik)$	$c_3(ik^+)$
1	527.2	773.7
2	508.7	630.7
3	671.4	752.0
4	967.2	1027.4
5	1394.1	

FIGURE 2

quation (4.5). We got for examples 1 and 4

Example	T_0^*	$C_S(T_0^*)$	T_1^*	$C(T_1^*)$
1	0.379	113.51	0.590	313.72
4	0.376	387.92	0.463	616.97



i	$c_3(ik)$	$c_3(ik^+)$
1	157.9	1152.3
2	613.9	1101.5
3	808.8	1127.6
4	967.2	1198.7
5	1196.3	

FIGURE 3

Figure 3 illustrates Example 1, but with $p=200$. Finally, it was interesting to compare the "age replacement policy" with and without adjustment costs, with the "periodic replacement policy with minimal repair," with and without adjustment costs. Let us consider Example 4 and denote by $A(T)$ and $A_S(T)$ the costs for the "age" problem with and without adjustment costs. Then, using the above results, and results from [1], we get:

$A_S(T_0^*)$	= 371.572	T_0^*	= 0.381	$C_S(T_0^*)$	= 387.92	T_0^*	= 0.376
$A(T^*)$	= 498.581	T^*	= 0.400	$C(T^*)$	= 508.69	T^*	= 0.400
$A(T_1^*)$	= 598.284	T_1^*	= 0.468	$C(T_1^*)$	= 616.97	T_1^*	= 0.463

It was seen from our computations that the two optimal policies (age policy and minimal repair at failure policy) do not differ very much from one another, and neither did the associated minimum costs.

REFERENCES

[1] Cléroux R. and M. Hanscom, "Age Replacement with Adjustment and Depreciation Costs and Interest Charges," Technical Report 145, Département d'Informatique, Université de Montréal (1973). To appear in Technometrics.

[2] Scheaffer, R. L., "Optimum Age Replacement Policies with an Increasing Cost Factor," Technometrics *13*, 139–144. (1971).

[3] Barlow, R. E. and F. Proschan, *Mathematical Theory of Reliability* (J. Wiley, 1965).

[4] Barlow, R. E. and L. C. Hunter, "Optimum Preventive Maintenance Policies", Operations Res. *8*, 90–100 (1960).

[5] Ross, S. M., *Applied Probability Models with Optimization Applications* (Holden Day, 1970).

OPTIMAL SCHEDULES FOR SPECIAL STRUCTURE FLOWSHOPS

Jatinder N. D. Gupta
U.S. Postal Service
Washington, D.C.

ABSTRACT

This paper considers the problem of scheduling a given number of jobs on a specified number of machines in a flowshop where the objective function is to minimize the total throughput time in which all jobs complete processing on all machines. Based on the combinatorial analysis of the problem, several simple algorithms are developed for solving special structure flowshop scheduling problems where the process times are not completely random, but bear a well-defined relationship to one another. The proposed algorithms are both simple and computationally efficient and can optimally solve large-sized problems even with manual computational devices.

INTRODUCTION

A frequently occurring operational problem is one of finding the order in which a number of jobs (commodities) can be processed on a given number of machines (facilities) to optimize a specified objective function. If the given number of jobs is to be processed on a specified number of machines in such a way that the technological order of all jobs on all machines is the same, then the problem is said to be a flowshop scheduling problem. The ordering of jobs need not be the same for each machine in a flowshop. If the order of processing of all jobs is constrained to be the same on each machine (no passing of jobs allowed), and if n is the number of jobs, then there are $n!$ possible job orderings, called schedules, for the flowshop scheduling problem. As the number of jobs increases, the number of possible schedules becomes enormously large and the solution by complete enumeration becomes uneconomical and unoperational. In order to make these larger flowshop scheduling problems solvable, the already restrictive case of flowshop must have further simplifying conditions imposed. The most common of these assumptions are the following [2–5]:

1. Process times of all jobs on all machines are known and are deterministic (no variations in process times are allowed).
2. Machines may process only one job at a time, and jobs once started must be processed to completion (no splitting of jobs is permitted).
3. Intermachine transportation and machine set-up time are either ignored or included in processing time, and are independent of the order in which jobs are processed.
4. All jobs are ready to begin processing at once, and all machines are initially idle, ready to process any job.
5. All jobs are assigned equal priorities and are independent of each other.
6. There is only one machine of each type.
7. No cancellation of jobs occurs after the jobs have been accepted for processing in the shop.
8. Jobs are processed as early as possible.

Most of these, unfortunately, depart in varying degrees from real-life situations, but, except in the smaller problems, they are necessary in order to keep the formulation and computation within bounds of feasibility.

The purpose of flowshop scheduling techniques is to determine the schedule that minimizes or maximizes some well-defined objective function. The most common measure of schedule performance is make-span, defined as the throughput time in which all jobs complete processing on all machines. While make-span is often an appropriate measure of schedule performance, it is by no means the only appropriate measure [1, 5, 6]. However, computational difficulties and the desirability of maximizing the production rate generally makes make-span the most acceptable schedule performance and is used as the objective of flowshop scheduling in this paper.

The recent advances in flowshop scheduling techniques have shown that it is rather difficult to develop exact optimization techniques for the solution of the general scheduling problem of the flowshop. As a result of this awareness, several authors have considered special structure flowshop problems. Thus, for example, Johnson [7] considered the special three-machine problem where an additional condition on the process times relationship of various jobs resulted in the development of a simple and efficient optimization algorithm for seeking the desired schedule. Similar special cases are analyzed by Nabeshima [8, 9], Smith [10], and Szwarc [12].

In this paper, these special cases are generalized and simple algorithms are developed for solving several special structure flowshop scheduling problems where the objective function is one of minimizing the make-span.

PROBLEM FORMULATION AND SOLUTION APPROACH

A simple definition of flowshop scheduling is:

"Given n jobs to be processed on M machines in the same order, the process time of job a on machine m being t_{am}; $(a \leq n;\ m \leq M)$, find the common order (schedule) in which these n jobs should be processed on the M machines so as to minimize a specified measure of schedule performance."

Consider a partial schedule σ and a job a, which, when augmented to σ, results in another partial schedule σa. Then, from the physical conditions and assumptions of the problem outlined above, the completion time of partial schedule σa at machine m,represented by $T(\sigma a, m)$, is given by the following recursive relation [6]:

$$T(\sigma a, m) = \max\ [T(\sigma, m);\quad T(\sigma a, m-1)] + t_{am}, \tag{1}$$

where

$$T(\sigma a, 0) = T(\Phi, m) = 0; \qquad \forall \sigma a \text{ and } m,$$

and Φ is a null (empty) schedule.

When the measure of schedule performance is make-span, the flowshop scheduling problem, as outlined above, is one of minimizing $T(\sigma a, M)$, where σ ranges over all the possible permutations of $(n-1)$ jobs not containing job a and job a ranges from 1 through n.

From the above original problem (henceforth referred to as P), we define a reverse problem P' such that, the process time of job a on machine m, t'_{am}, is given by

$$t'_{am} = t_{a,M-m+1}, \qquad \forall m \leq M. \tag{2}$$

Then, the following results establish that the schedule which minimizes make-span for P is the reverse of the schedule which minimizes make-span for problem P'. Analogous to (1) above, we define

$$L(\sigma a, m) = \max\ [L(\sigma, m);\quad L(\sigma a, m-1)] + t'_{am}, \tag{3}$$

where

$$L(\sigma a, 0) = L(\Phi, m) = 0, \forall m \text{ and } \sigma a.$$

LEMMA 1: For any two schedules $\sigma = (a_1, a_2, \ldots, a_k)$ and $\sigma' = (a_k, a_{k-1}, \ldots, a_1)$:

$$T(\sigma, M) = L(\sigma', M). \tag{4}$$

PROOF: By repeated use of relation (3) above, we get

$$L(\sigma', M) = \max_{1 \le m'_{i-1} \le m'_i \le M} \left[\sum_{i=1}^{k} \sum_{s=m'_{i-1}}^{m'_i} t' a_{k-i+1} s \right]. \tag{5}$$

Without loss of generality, we can let

$$M - m'_i + 1 = m_{k-i}.$$

Then, with the help of relation (2) above, we can rewrite Equation (5) as

$$L(\sigma', M) = \max_{1 \le m_{i-1} \le m_i \le M} \left[\sum_{i=1}^{k} \sum_{s=m_{i-1}}^{m_i} t_{a_i s} \right]. \tag{6}$$

Therefore, (4) follows from (6) and the repeated use of relation (1). Hence the validity of Lemma 1.

THEOREM 1: If $S = (a_1, a_2, \ldots, a_n)$ is an optimal schedule for problem P, then, $S' = (b_n, \ldots, a_1)$ is an optimal schedule for problem P'.

PROOF: Let the complete set of schedules be $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$. Then, for any $\pi_i \epsilon \pi$, it follows that:

$$T(S, M) \le T(\pi_i, M)$$

Then, from Lemma 1, we get:

$$L(S', M) \le L(\pi'_i, M) \qquad \forall \pi'_i \epsilon \pi'.$$

Therefore, schedule S' minimizes make-span for problem P' and hence is optimal, thus proving the validity of Theorem 1.

Theorem 1 above provides a means of two alternative solution procedures for solving P or P'. Thus, depending on the specific situation, we may prefer to solve P' instead of P (or vice-versa). In

order to solve the problem P, consider two schedules $\sigma ab\pi$ and $\sigma ba\pi$ where $\sigma \not\ni a, b, \pi$; $a, b \not\in \pi$ and $a \neq b$. Based on the analysis of these two schedules, the combinatorial analysis approach, originally outlined by Wayne Smith [11] and subsequently used by several authors (see References [3] and [4] for recent reviews), attempts are made to derive sufficient conditions to obtain an optimal solution and prove its optimality. The efficiency of the combinatorial algorithm thus obtained depends on the number and type of optimality conditions derived. Until this date, Wayne Smith's theorem remains classic and provides an easiest and simplest approach to sequencing and scheduling problems. We state a simplified version of his theorem and show that the approach is applicable to several practical flowshop scheduling problems.

THEOREM 2: If for each $i \leq n$, there exists a function $f(i) = f(t_{i1}, t_{i2}, \ldots, t_{iM})$ such that $f(a) \leq f(b)$ implies that $T(\sigma ab\pi, M) \leq T(\sigma ba\pi, M), \forall \sigma$ and π, then a schedule obtained by arranging jobs in ascending order of $f(i)$ minimizes make-span.

PROOF: Let the schedule obtained by arranging the jobs in ascending order of $f(i)$ be S. Consider another schedule S' such that if $S = \sigma a\pi'\pi''$, then $S' = \sigma\pi' a\pi''$. Since the schedule S is obtained by arranging the jobs in ascending order of $f(i)$, it follows that

$$f(a) \leq f(b), \qquad \forall b \epsilon \pi'. \tag{7}$$

For notational convenience, let $\pi' = (b_1, b_2, \ldots, b_k)$ and

$$\eta_i = (b_1, b_2, \ldots, b_i), \qquad \eta_i' = (b_{i+1}, \ldots, b_k), \qquad \forall i \leq k$$

$$\eta_0 = \eta_k' = \Phi \qquad \text{and} \qquad \eta_k = \eta_0' = \pi'.$$

From the condition of Theorem 2 and relation (7) it follows that

$$T(\sigma\eta_{k-1} a\eta_{k-1}' \pi'', M) \leq T(\sigma\pi' a\pi'', M),$$

since $f(a) \leq f(b_k)$.

Similarly,

$$T(\sigma\eta_{k-2} a\eta'_{k-2}\pi'', M) \leq T(\sigma\eta_{k-1} a\eta'_{k-1}\pi'', M) \leq T(\sigma\pi' a\pi'', M).$$

Proceeding recursively as above, we observe that

$$T(\sigma a\pi'\pi'', M) \leq T(\sigma\pi' a\pi'', M).$$

Since σ, π', and π'' are of arbitrary length (contain arbitrary number of jobs), it follows that S has a make-span less than or equal to the make-span of any other schedule S'. Hence the schedule obtained by arranging the jobs in ascending order of $f(i)$ is a minimum make-span schedule and thus optimal.

Let $p(i) = p(t_{i1}, t_{i2}, \ldots, t_{iM})$ and $q(i) = w(t_{i1}, t_{i2}, \ldots, t_{iM})$ be two non-negative real functions of the process times of job i, $\forall i \leq n$. Define

8) $$f(i) = \frac{\text{sign } (p(i) - q(i))}{\min (p(i), q(i))}$$

where sign $(p(i) - q(i))$ is the algebraic sign of the difference $p(i) - q(i)$ and equals 1 if $p(i) \geq q(i)$ and -1 if $p(i) < q(i)$. Then, for any job pair (a, b), the following Lemma results:

LEMMA 2: If

(9) $$f(a) \leq f(b), \text{ then}$$

(10) $$\min [p(a), q(b)] \leq \min [p(b), q(a)]$$

PROOF: We consider all the three possible cases in which the condition of Lemma 2 holds:

(i) $p(a) < q(a)$; $p(b) < q(b)$ in which case (9) implies that $p(a) \leq p(b)$.
(ii) $p(a) < q(a)$; $p(b) \geq q(b)$
(iii) $p(a) \geq q(a)$; $p(b) \geq q(b)$, in which case (9) implies that $q(b) \leq q(a)$.

In any of the above three cases, it is obvious that (10) holds. Therefore, (9) implies (10) and hence he proof of Lemma 2.

Lemma 2 represents a simplification of Johnson's two-machine condition and clearly shows the transitivity of the conditions. We will use this Lemma 2 for stating the results of the special cases considered subsequently. However, the essential difficulty arises because, for the general M-stage flowshop scheduling problem, the function $f(i)$ cannot be identified and hence Theorem 2 cannot be implemented [3]. The special cases considered in this paper are such that the function $f(i)$ can be obtained for varying schedules and hence simple algorithms developed.

SOLUTION OF SPECIAL STRUCTURE FLOWSHOP PROBLEMS

The special structure flowshop scheduling problems considered here are the results of the belief that the common technological order of jobs and the "no-passing allowed" assumption in a shop does reflect certain similarities in the process times of jobs on different machines. However, without knowing the actual shop conditions, it is difficult to predict the exact nature of the relationships between several process times. Since different situations may give rise to different relationships, we describe simple algorithms for finding optimal schedules for five special cases of the general flowshop scheduling problems.

CASE 1: For some one fixed $h \leq (M-1)$, the process times of jobs satisfy the following conditions:

$$\text{(i)} \quad \min_{1 \leq i \leq n} t_{im} \geq \max_{1 \leq j \leq n} t_{jm+1}, \qquad \forall\, m \leq h-1 \leq M-2$$

and

11) $$\text{(ii)} \quad \max_{1 \leq i \leq n} t_{im} \leq \min_{1 \leq j \leq n} t_{jm+1}, \forall\, h+1 \leq m \leq M-1.$$

When $M=2$, the conditions (11) are not required since the possible value of $h=1$. However, when $M=3$, either (i) or (ii) of the above conditions is active and represents Johnson's special three-

machine cases. Hence, case 1 is the extension of Johnson's special three-machine case to general M-machine case and has been studied by Nabeshima [8, 9]. The following results develop the algorithm and establish its proof of optimality.

LEMMA 3: If conditions (11) are satisfied; then for any schedule σ and job a:

$$T(\sigma, m) \leq T(\sigma, 1) + \sum_{s=1}^{m-1} t_{as}, \qquad \forall\, m \leq h \leq M-1. \tag{12}$$

PROOF: If σ is empty, then (12) directly follows from (1). However, if $\sigma \neq \Phi$, then, let $\sigma = (a_1, a_2, \ldots, a_k)$. Repeated use of relation (1) yields

$$T(\sigma a, m) = \max_{1 \leq m_1 \leq m_2 \leq \ldots \leq m_k \leq m} \left[\sum_{s=1}^{m_1} t_{a_1 s} + \sum_{s=m_1}^{m_2} t_{a_2 s} + \ldots + \sum_{s=m_{k-1}}^{m_k} t_{a_k s} + \sum_{s=m_k}^{m} t_{as} \right].$$

The condition (11) of Lemma 3 implies that

$$\sum_{s=m_{i-1}}^{m_i} t_{a_i s} + \sum_{s=m_i}^{m_{i+1}} t_{a_{i+1} s} \leq t_{a_i m_{i-1}} + \sum_{s=m_{i-1}}^{m_{i+1}} t_{a_{i+1} s}, \qquad \forall i \leq k-1 \text{ and } \forall\, m \leq h,$$

hence

$$T(\sigma a, m) = \sum_{i=1}^{k} t_{a_i 1} + \sum_{s=1}^{m} t_{as}, \qquad \forall m \leq h. \tag{13}$$

Since $T(\sigma, 1) = \sum_{i=1}^{k} t_{a_i 1}$ and $T(\sigma a, m) \geq T(\sigma, m) + t_{am}$, (12) follows from (13) thus proving Lemma 3.

LEMMA 4: If conditions (11) are satisfied, then, for any nonempty schedule σ, and job a

$$T(\sigma, M) \geq T(\sigma, m) + \sum_{s=m}^{M-1} t_{as}, \forall\, h+1 \leq m \leq M-1. \tag{14}$$

PROOF: Let the last job of σ be c. Then, from relation (1) above it follows that

$$T(\sigma, s) - T(\sigma, s-1) \geq t_{cs}, \qquad \forall\, s \leq M,$$

with condition (11), therefore,

$$T(\sigma, s) - T(\sigma, s-1) \geq t_{as-1}, \qquad \forall h+1 \leq s-1 \leq (M-1). \tag{15}$$

The summation of expression (15) over the range $s = m+1$ through $s = M$ results in (14) above, thus proving Lemma 4.

LEMMA 5: Consider a schedule σa for the problem satisfying condition (11). Then

$$T(\sigma a, M) = \max \left[T(\sigma, 1) + \sum_{s=1}^{M} t_{as}; \quad T(\sigma, M) + t_{aM} \right]. \tag{16}$$

PROOF: If $\sigma = \Phi$ (empty), then (16) directly follows from relation (1) above. However, if $\sigma \neq \Phi$ then, the repeated use of relation (1) gives

$$T(\sigma a, M) = \max_{1 \leq m \leq M} \left[T(\sigma, m) + \sum_{s=m}^{M} t_{as} \right], \tag{17}$$

$$\forall m \leq h, \quad T(\sigma, m) + \sum_{s=m}^{M} t_{as} \leq T(\sigma, 1) + \sum_{s=1}^{M} t_{as}, \text{ (Lemma 3), and} \tag{18a}$$

$$\forall m > h, \quad T(\sigma, m) + \sum_{s=m}^{M} t_{as} \leq T(\sigma, M) + t_{aM} \text{ (Lemma 4).} \tag{18b}$$

Hence (16) follows from (17) and (18) thus establishing the validity of Lemma 5.

LEMMA 6: Consider two schedules σab and σba. If the process times satisfy conditions (11) and

$$\min \left[\sum_{m=1}^{M-1} t_{am}; \sum_{m=2}^{M} t_{bm} \right] \leq \min \left[\sum_{m=1}^{M-1} t_{bm}; \sum_{m=2}^{M} t_{am} \right], \tag{19}$$

then,

$$T(\sigma ab, M) \leq T(\sigma ba, M). \tag{20}$$

PROOF: Repeated use of Lemma 4 above shows that

$$T(\sigma ab, M) = \max \left[T(\sigma, 1) + t_{a1} + \sum_{m=1}^{M} t_{bm}; \quad T(\sigma, 1) + \sum_{m=1}^{M} t_{am} + t_{bM}; \right. \tag{21}$$

Similarly

$$\left. T(\sigma, M) + t_{aM} + t_{bM} \right].$$

$$T(\sigma ba, M) = \max \left[T(\sigma, 1) + t_{b1} + \sum_{m=1}^{M} t_{am}; \quad T(\sigma, 1) + \sum_{m=1}^{M} t_{bm} + t_{aM}; \right. \tag{22}$$

$$\left. T(\sigma, M) + t_{bM} + t_{aM} \right].$$

Subtracting $T(\sigma, 1) + \sum_{m=1}^{M} (t_{am} + t_{bm})$ from both sides of (19) and multiplying the resulting expressions by (-1), we get

$$\max \left[T(\sigma, 1) + t_{a1} + \sum_{m=1}^{M} t_{bm}; \quad T(\sigma, 1) + \sum_{m=1}^{M} t_{am} + t_{bM} \right] \leq \max \left[T(\sigma, 1) + t_{b1} + \sum_{m=1}^{M} t_{am}; \right. \tag{23}$$

$$\left. T(\sigma, 1) + \sum_{m=1}^{M} t_{bm} + t_{aM} \right].$$

Since $T(\sigma, M)+t_{aM}+t_{bM}$ is common to both (21) and (22), (20) follows from (21), (22), and (23), thus the proof of Lemma 6.

We define a function:

$$f(i)=\frac{\text{Sign } (t_{i1}-t_{iM})}{\min\left[\sum_{m=1}^{M-1} t_{im};\ \sum_{m=2}^{M} t_{im}\right]}, \qquad \forall i \leqslant n. \tag{24}$$

Then, the following theorem gives the algorithm to find optimal schedule for Case 1.

THEOREM 3: For a flowshop scheduling problem satisfying conditions (11), an optimal schedule is obtained by arranging the jobs in ascending order of $f(i)$ given by Equation (24).

PROOF: It is sufficient to show that, for any two schedules $\sigma ab\pi$ and $\sigma ba\pi$, $f(a) \leqslant f(b)$ implies that $T(\sigma ab\pi, M) \leqslant T(\sigma ba\pi, M)$ since the proof then follows from Theorem 2. From Lemma 2, the condition $f(a) \leqslant f(b)$ implies condition (19) of Lemma 6. Hence, from Lemma 6

$$T(\sigma ab, M) \leqslant T(\sigma ba, M).$$

Let $\pi = c\eta$. From Lemma 5

$$T(\sigma abc, M) = \max\left[T(\sigma ab, 1)+\sum_{s=1}^{M} t_{cs};\quad T(\sigma ab, M)+t_{cM}\right].$$

Since $T(\sigma ab, 1) = T(\sigma ba, 1) = \sum_{i \epsilon \sigma ab} t_{i1}$ and $T(\sigma ab, M) \leqslant T(\sigma ba, M)$, it follows that

$$T(\sigma abc, M) \leqslant \max\left[T(\sigma ba, 1)+\sum_{s=1}^{M} t_{cs};\quad T(\sigma ba, M)+t_{cM}\right] = T(\sigma bac, M).$$

Similarly by letting $\eta = d\eta'$, we can show that

$$T(\sigma abcd, M) \leqslant T(\sigma bacd, M).$$

This argument, when repeated recursively results in $T(\sigma ab\pi, M) \leqslant T(\sigma ba\pi, M)$. Therefore, the proof of Theorem 3 follows from Theorem 2.

CASE 2: The process times of this case satisfy the following condition:

$$\max_{1 \leqslant i \leqslant n} t_{im} \leqslant \min_{1 \leqslant j \leqslant n} t_{jm+1}, \qquad \forall m \leqslant (M-2). \tag{25}$$

LEMMA 7: If the process times satisfy conditions (25), then, for any nonempty schedule σ and job a

$$T(\sigma a, M) = \max\left[T(\sigma, M-1)+t_{aM-1}+t_{aM};\ T(\sigma, M)+t_{aM}\right]. \tag{26}$$

PROOF: From Lemma 4 (as a special case when $M = M-1$),

$$T(\sigma, M-1) \geq T(\sigma, m) + \sum_{s=m}^{M-2} t_{as}, \qquad \forall m \leq (M-2). \tag{27}$$

From relation (1) above

$$T(\sigma a, M) = \max_{1 \leq m \leq M} \left[T(\sigma, m) + \sum_{s=m}^{M} t_{as} \right]. \tag{28}$$

Therefore (26) follows from (27) and (28), hence the proof of Lemma 7. Define

$$g(i) = \text{Sign } (t_{iM-1} - t_{iM}) / \min (t_{iM-1}; t_{iM}), \qquad \forall i \leq n. \tag{29}$$

Then the following theorem immediately results.

THEOREM 4: For the flowshop scheduling problem satisfying conditions (25) and any fixed initial starting job, a minimum make-span schedule is obtained by arranging the remaining $(n-1)$ jobs in ascending order of $g(i)$ given by Equation (29).

PROOF: We only briefly outline the proof of this theorem since it closely follows the developments of Case 1, except that in Case 2, σ cannot be empty. Let the fixed initial job be a_1. Consider two jobs a and b such that $g(a) \leq g(b)$. From Lemma 2 and analysis similar to Lemma 6 yields that $T(\sigma ab, M) \leq T(\sigma ba, M)$, $\forall \sigma \neq \Phi$. Then, following the analysis similar to that of Theorem 3 shows that if the conditions (25) are satisfied, then $g(a) \leq g(b)$ implies $T(\sigma ab\pi, M) \leq T(\sigma ba\pi, M)$. Let $\sigma = a_1\phi$. Then $T(a_1\phi ab\pi, M) \leq T(a_1\phi ab\pi, M$. Hence with fixed a_1, a minimum make-span schedule is obtained by arranging the remaining $(n-1)$ jobs in ascending order of $g(i)$.

Theorem 4 above provides a method to solve Case 2. For each $i \leq n$, we can generate n schedules according to Theorem 4 and the minimum make-span schedule among these n schedules is optimal. However, we can reduce the number of schedules by obtaining the first job(s) according to the following theorem:

THEOREM 5: Consider two jobs a and b. If conditions (25) hold and

$$\left.\begin{aligned} & g(a) \leq g(b) \\ & \sum_{s=1}^{M-2} t_{as} \leq \sum_{s=1}^{M-2} t_{bs} \end{aligned}\right\}, \tag{30}$$

then

$$T(ab\pi, M) \leq T(ba\pi, M).$$

PROOF: From Lemma 7 above and conditions (30)

$$T(ab, M-1) = T(a, M-1) + t_{bM-1} = \sum_{s=1}^{M-1} t_{as} + t_{bM-1} \leq \sum_{s=1}^{M-1} t_{bs} + t_{aM-1} = T(ba, M-1). \tag{31}$$

$$T(ab, M) = \max\,[T(ab, M-1); T(a, M)] + t_{bM} \tag{32}$$

$$= \max\left[\sum_{s=1}^{M-1} t_{as} + t_{bM-1}; \sum_{s=1}^{M} t_{as}\right] + t_{bm}$$

$$= \max\,[-t_{aM}; -t_{bM-1}] + \sum_{s=1}^{M} t_{as} + t_{bM-1} + t_{bM}.$$

Similarly,

$$T(ba, M) = \max\,[-t_{bM}; -t_{aM-1}] + \sum_{s=1}^{M} t_{bs} + t_{aM-1} + t_{aM}. \tag{33}$$

From conditions (30) and Lemma 2 it follows that

$$\max\,[-t_{bM-1}; -t_{aM}] + \sum_{s=1}^{M-2} t_{as} \leq \max\,[-t_{aM-1}; -t_{bM}] + \sum_{s=1}^{M-2} t_{bs}, \tag{34}$$

hence (32), (33), and (34) lead to

$$T(ab, M) \leq T(ba, M). \tag{35}$$

Now let $\pi = c\eta$. Then, from Lemma 7

$$T(abc, M) = \max\,[T(ab, M-1) + t_{cM-1} + t_{cM}; T(ab, M) + t_{cM}],$$

and from (31) and (35)

$$T(abc, M) \leq \max\,[T(ba, M-1) + t_{cM-1} + t_{cM}; T(ba, M) + t_{cM}] = T(bac, M).$$

The remaining proof follows by recursion of $\eta = d\eta'$ etc. Hence the job a satisfying conditions (30) $\forall b \leq n$ is our starting job. However, the case is somewhat difficult since we have an additional condition here and no single job may satisfy both the conditions. Thus, we may have to consider more than one job at the starting position.

The step by step procedure to solve problems of Case 2 is as follows:

STEP 1: For each $i \leq n$, find $g(i) = \text{Sign}\,(t_{iM-1} - t_{iM})/\min\,(t_{iM-1}, t_{iM})$ and $X_i = \sum_{s=1}^{M-2} t_{is}$.

STEP 2: Arrange the jobs in ascending order of $g(i)$. Let the schedule so obtained be $S = (a_1, a_2, \ldots, a_n)$.

STEP 3: Let $\pi = (b_1, b_2, \ldots, b_k)$ be the set of jobs such that

$$X_{b_i} \leq X_{a_1}, \qquad \forall\ i \leq k; \qquad b_i \neq a_1.$$

Generate k more schedules by assigning b_i at the first sequence position and maintaining schedule S order for remaining jobs.

STEP 4: For the $(k+1)$ schedules so obtained, find the make-spans. The schedule with minimum make-span is optimal.

EXAMPLE: As an illustration of Case 2, consider the following 4-job, 3-machine problem:

i \ m	1	2	3	$g(i)$	X_i
1	8	12	3	1/3	8
2	7	9	5	1/5	7
3	5	8	8	1/8	5
4	7	10	12	−1/10	7

Arranging the jobs in ascending order of $g(i)$, we get the schedule (4, 3, 2, 1); however, $X_4 \nleq X_3$, thus $\pi = (3)$. We obtain two schedules (4, 3, 2, 1), and (3, 4, 2, 1) with make-spans 49 and 47. Schedule (3, 4, 2, 1), therefore, is optimal and optimal (minimum) make-span is 47.

CASE 3: The process times of this case satisfy the following conditions:

$$\max_{1\leq i\leq n} t_{im+1} \leq \min_{1\leq j\leq n} t_{jm}, \qquad \forall\ 2 \leq m \leq M-1. \tag{36}$$

We solve the reverse problem P' for Case 3. From relation (2), it follows that

$$\max_{1\leq j\leq n} t'_{jm} \leq \min_{1\leq i\leq n} t'_{im+1}, \qquad \forall m \leq M-2. \tag{37}$$

Since conditions (37) are the same as (25), we can find optimal schedule for P' by using the algorithm for Case 2. Reversing the schedule thus obtained yields optimal schedule for problem P satisfying conditions (36). Thus, the algorithm for Case 3 is as follows:

STEP 1: For each $i \leq n$, calculate $f(i) = \text{Sign}\ (t_{i1} - t_{i2})/\min\ (t_{i1}, t_{i2})$ and $X_i = \sum_{s=3}^{M} t_{is}$.

STEP 2: Arrange the jobs in ascending order of $f(i)$. Let the schedule so generated be $S = (a_1, a_2, \ldots, a_k, \ldots, a_n)$.

STEP 3: Let $\pi = (b_1, b_2, \ldots, b_k)$ be set such that $X_{b_i} \leq X_{a_n}$, $i \leq k$, $b_i \neq a_n$. Generate k more schedules by assigning job b_i at the last sequence position and maintaining schedule S order for remaining jobs.

STEP 4: Compute the make-span of the $(k+1)$ schedules so generated. The schedule with minimum make-span is optimal.

CASE 4: In this special structure problem $t_{im} \leq t_{iM}, \forall i \leq n$ and $\forall m \leq M$. In addition, the process times are such that there exists a schedule $S = (a_1, a_2, \ldots, a_k, \ldots, a_n)$ such that $\forall i \leq n-1$,

$t_{a_i m} \leq t_{a_{i+1} m}$, $\forall m \leq M$. Further, either $t_{im} \leq t_{jm}$ or $t_{im} \geq t_{jm}$ $\forall i, j \leq n$. The following developments prove that schedule S obtained above minimizes make-span.

LEMMA 8: If $t_{am} \leq t_{bm}, \forall m \leq M$, then, for two schedules σa and σb:

$$T(\sigma a, m) - T(\sigma b, m) \leq t_{am} - t_{bm} \leq 0. \tag{38}$$

PROOF: From recursive relation (1) above, we get

$$T(\sigma a, m) - T(\sigma b, m) = \max_{1 \leq u \leq m} \left[T(\sigma, u) + \sum_{s=u}^{m} t_{as} \right] - \max_{1 \leq u \leq m} \left[T(\sigma, u) + \sum_{s=u}^{m} t_{bs} \right] \tag{39}$$

to

$$\leq \max_{1 \leq u \leq m} \left[\sum_{s=u}^{m} (t_{as} - t_{bs}) \right],$$

since $t_{as} \leq t_{bs}$ $\forall s \leq M$, it follows that

$$\sum_{s=u}^{m} (t_{as} - t_{bs}) \leq t_{am} - t_{bm} \leq 0, \qquad \forall m \leq M. \tag{40}$$

Therefore (38) follows from (39) and (40), thus proving Lemma 8.

LEMMA 9: If $t_{am} \leq \min [t_{aM}, t_{bm}]$ $\forall m \leq M$, then for any two schedules σab and σba:

$$T(\sigma ab, M) \leq T(\sigma ba, M). \tag{41}$$

PROOF: By repeated use of relation (1) above

$$T(\sigma ab, M) = \max_{1 \leq m \leq u \leq M} \left[T(\sigma, m) + \sum_{s=m}^{u} t_{as} + \sum_{s=u}^{M} t_{bs} \right]$$

and with the condition of Lemma 9

$$T(\sigma ab, M) = \max_{1 \leq m \leq M} \left[T(\sigma, m) + t_{am} + \sum_{s=m}^{M} t_{bs} \right].$$

Since $t_{am} \leq t_{aM}$

$$T(\sigma ab, M) \leq \max_{1 \leq m \leq M} \left[T(\sigma, m) + \sum_{s=m}^{M} t_{bs} + t_{aM} \right]. \tag{42}$$

Similarly:

$$T(\sigma ba, M) = \max_{1 \leq m \leq u \leq M} \left[T(\sigma, m) + \sum_{s=m}^{u} t_{bs} + \sum_{s=u}^{M} t_{as} \right] \tag{43}$$

$$= \max_{1 \leq m \leq M} \left[T(\sigma, m) + \sum_{s=m}^{M} t_{bs} + t_{aM} \right].$$

Therefore (41) follows from (42) and (43), thus proving Lemma 9.

THEOREM 6: For Case 4 problems, the schedule $S = (a_1, a_2, \ldots, a_k, \ldots, a_n)$ minimizes make-span and hence is optimal.

PROOF: It is sufficient to show that $t_{am} \leq \min [t_{aM}; t_{bm}]$ imply that $T(\sigma ab\pi, M) \leq T(\sigma ba\pi, M)$ since, the proof then follows from theorem 2.

Let $\pi = c\eta$. Then

$$T(\sigma abc, M) = \max_{1 \leq m \leq M} \left[T(\sigma ab, m) + \sum_{s=m}^{M} t_{cs} \right]$$

$$= \max_{1 \leq m \leq u \leq M} \left[T(\sigma a, m) + \sum_{s=m}^{u} t_{bs} + \sum_{s=u}^{M} t_{cs} \right].$$

According to the conditions of the theorem, either $t_{cm} \leq t_{bm} \; \forall \; m$ or $t_{bm} \leq t_{cm} \; \forall \; m \leq M$.

Therefore,

(44) $$T(\sigma abc, M) = \max_{1 \leq m \leq M} \left[T(\sigma a, m) + t_{bm} + \sum_{s=m}^{M} t_{cs}; \; T(\sigma a, m) + \sum_{s=m}^{M} t_{bs} + t_{cM} \right]$$

$$= \max_{1 \leq m \leq M} \left[T(\sigma a, m) + t_{bm} + \sum_{s=m}^{M} t_{cs}; \; T(\sigma ab, M) + t_{cM} \right].$$

Similarly;

(45) $$T(\sigma bac, M) = \max_{1 \leq m \leq M} \left[T(\sigma b, m) + t_{am} + \sum_{s=m}^{M} t_{cs}; \; T(\sigma ba, M) + t_{cM} \right]$$

from (44) and Lemmas 8 and 9

(46) $$T(\sigma abc, M) \leq \max_{1 \leq m \leq M} \left[T(\sigma b, m) + t_{am} + \sum_{s=m}^{M} t_{cs}; \; T(\sigma ba, M) + t_{cM} \right].$$

Therefore (45) and (46) yield

$$T(\sigma abc, M) \leq T(\sigma bac, M).$$

Thus by recursive expansion of $\eta = d\eta'$ etc., it is seen that

$$T(\sigma ab\pi, M) \leq T(\sigma ba\pi, M).$$

Hence the schedule S is an optimal schedule since it minimizes make-span.

CASE 5: In this special case $t_{im} \leq t_{i1}, \forall i \leq n, \forall m \leq M$. In addition, the process times are such that there exists a schedule $S=(a_1, a_2, \ldots, a_n)$ such that $\forall i \leq n-1$, $t_{a_i m} \geq t_{a_{i+1} m}, \forall m \leq M$ and either $t_{im} \geq t_{jm}$ or $t_{jm} \leq t_{im}, \forall i, j \leq n$, and $\forall m \leq M$.

THEOREM 7: For Case 5, schedule S has a minimum make-span and hence is optimal.

PROOF: The conditions of Case 5 imply that $\forall i \leq n, t'_{im} \leq t'_{iM}, \forall m \leq M$ and there exists a schedule $S'=(a_n, a_{n-1}, \ldots, a_1)$ which satisfies the conditions of Case 4. Hence, by Theorem 6, S' is an optimal schedule for problem P'. Therefore, from Theorem 1 it follows that S is optimal for problem P satisfying Case 5.

The conditions of Cases 4 and 5 seem rather too restrictive. However, there are a series of practical flowshop scheduling problems that satisfy these conditions. Recently, Milton Smith [10] considered flowshop scheduling problems where

$$t_{im} = \frac{t_{i1} \cdot t_{1m}}{t_{11}}, \qquad \forall i \leq n, \qquad \forall m \leq M. \tag{47}$$

Smith called such problems as ordered matrix problems since there is an order maintained in the process times of various jobs on various machines. Thus if the process times of job 1 on all machines and of all jobs on machine 1 are known, other process times are determined by Equation (47) above. Among these ordered problems, Smith considers two special cases (a) $t_{1m} \leq t_{11} \leq t_{1M}, \forall m \leq M$ and (b) $t_{1m} \leq t_{1M} \leq t_{11}, \forall m \leq M$. It is interesting to note that the algorithms for Cases 4 and 5 above prove the specific cases considered by Smith and establish the optimality of his algorithm. Further, the algorithms proposed by Smith are applicable to a wider class of problems than envisioned by Smith. We state this as the following theorem:

THEOREM 8: Consider an ordered flowshop scheduling problem given by Equation (47). (a) If $t_{1m} \leq t_{1M}, \forall m \leq M$, an optimal schedule is obtained by arranging the jobs in ascending order of t_{i1}. (b) If $t_{1m} \leq t_{11}, \forall m \leq M$, then an optimal schedule is obtained by arranging the jobs in descending order of t_{i1}

PROOF: (a) From (47), $t_{1m} \leq t_{1M}$ implies that $t_{im} \leq t_{iM}, \forall m \leq M$ and $\forall i \leq n$. Let the schedule obtained by arranging the jobs in ascending order of t_{i1} be $S=(a_1, a_2, \ldots, a_n)$. Then, $\forall i \leq n-1$,

$$t_{a_i m} = \frac{t_{a_i 1} t_{1m}}{t_{11}} \tag{48}$$

and

$$t_{a_{i+1} m} = \frac{t_{a_{i+1} 1} t_{11}}{t_{11}}. \tag{49}$$

Since $t_{a_i 1} \leq t_{a_{i+1} 1}$, it follows that $t_{a_i m} \leq t_{a_{i+1} m}, \forall m \leq M$. Hence, from Theorem 6, the schedule S is optimal.

(b) From Equation (47), $t_{1m} \leq t_{11}$ implies that $t_{im} \leq t_{i1}, \forall i \leq n$, $m \leq M$. Let the schedule obtained by arranging the jobs in descending order of t_{i1} be $S=(a_1, a_2, \ldots, a_n)$. Then: $\forall i \leq (n-1)$, (48), and (49) imply that $t_{a_i m} \geq t_{a_{i+1} m}$ since $t_{a_i 1} \geq t_{a_{i+1} 1}$. Hence the conditions of Case 5 are satisfied and, from

Theorem 7, the schedule S is optimal.

The results of Theorem 8 above also show that the additional restriction $t_{im} \leq t_{i1} \leq t_{iM}\ \forall i \leq n$ and $\forall m \leq M$ for Case 4 and $t_{im} \leq t_{iM} \leq t_{i1}, \forall i \leq n$, and $\forall m \leq M$ for Case 5 problems as originally conceived by Smith [10] are unnecessary and are not required.

CONCLUSIONS

We have discussed five special cases of flowshop scheduling problems and have developed simple algorithmic rules for solving these problems when the objective function is minimization of make-span and no passing of jobs is allowed. The problem analysis and algorithm developments show that a relationship between various process times actually results in a simpler solution to the problem than completely random case. However, the five special cases considered here do not exhaust the possibility of all special cases that may occur in practice. It is my feeling that future developments be directed towards finding new special cases and simple algorithmic rules to solve such problems since the approach then can be used to secure an efficient technique to solve the general flowshop scheduling problem as a conglomeration of special cases.

REFERENCES

[1] Conway, R. W., W. L. Maxwell, and L. W. Miller, *Theory of Scheduling* (Addison Wesley Publishing Co., Readings, Mass., 1967).

[2] Dudek, R. A., and O. F. Teuton, Jr., "Development of M–Stage Decision Rule for Scheduling n Jobs through m Machines," Operations Research, *12*, 471–497 (1964).

[3] Elmaghraby, S. E., "The Machine Sequencing Problem—Review and Extensions," Nav. Res. Log. Quart. *15*, 205–232 (1968).

[4] Gupta, J. N. D., "M–Stage Scheduling Problem—A Critical Appraisal," The International Journal of Production Research *9*, 267–281 (1971).

[5] Gupta, J. N. D., "Economic Aspects of Production Scheduling Systems," Journal of Operations Research Society of Japan *13*, 167–192 (1971).

[6] Gupta, J. N. D. and R. A. Dudek, "Optimality Criteria for Flowshop Schedules," AIIE Transactions *3*, 199–206 (1971).

[7] Johnson, S. M., "Optimal Two and Three-Stage Production Schedules with Setup Times Included," Nav. Res. Log. Quart. *1*, 61–68 (1954).

[8] Nabeshima, I., "The Order of n Items Processed on m Machines," Journal of Operations Research Society of Japan *3*, 170–75 (1961).

[9] Nabeshima, I., "Some Extensions of the M–Machine Scheduling Problem," Journal of Operations Research Society of Japan *10*, 1–17 (1967).

[10] Smith, M. L., "A Critical Analysis of Flowshop Sequencing," Ph. D. Dissertation, Texas Technological College, Lubbock, Texas (1968).

[11] Smith, W. E., "Various Optimizers for Single-Stage Production," Nav. Res. Log. Quart. *3*, 56–66 (1956).

[12] Szwarc, W., "On Some Sequencing Problems," Nav. Res. Log. Quart. *15*, 127–155 (1968).

SENSITIVITY OF QUEUEING MODELS TO THE ASSUMPTION OF EXPONENTIALITY*

Donald Gross

The George Washington University
Washington, D.C.

ABSTRACT

This study deals with determining the effects on queuing models of assuming that interarrival and service times are exponentially distributed. Steady-state expected value measures of congestion are compared for a variety of GI/G models when assuming an M/M model as an approximation. The majority of this effort involves single-channel queues; however, some multichannel situations are also considered. In addition, effects of the exponential approximation are observed in certain cost-optimization models.

1. INTRODUCTION

Applied analytical work in queueing theory generally relies on the assumptions of exponentiality for interarrival or service times in order to use the very tractable M/M queuing formulas. Often, these assumptions are not met in practice. In a classic study by Brigham [1], an $M/M/c$ model was assumed after visual inspection of arrival and service data. Upon subjecting these data to χ^2 goodness-of-fit tests, there was evidence that both service and arrival times may be nonexponential (see chap. 9 of [2]). This, then, poses a dilemma to the analyst. He has essentially three choices: (1) use M/M as an approximation; (2) use direct approximations for GI/G, such as bounds, inequalities, or heavy traffic estimators; or (3) use Monte Carlo simulation. This paper looks in detail at the first choice. There has been little attention paid this approach in the literature, although P. M. Morse (chaps. 7 and 8 of [6]) hints at this problem. Quite a lot of attention has been devoted to the second approach [3], [4], [5] as well as the third (see, for example, chap. 8 of [2]).

For this study, the criterion of effectiveness of the M/M as an approximating model in non-M/M cases is the percentage error incurred when using the approximation. Percentage errors in expected-value measures of congestion and in optimal costs in some cost-optimization models are considered.

The expected-value measures of congestion of interest are generally the mean system size, L, the mean queue size, L_q, the mean queue wait, W_q, and the mean system wait, W. Because of Little's formula which relates these measures, it is necessary to consider only one of them. If, for example, we decide to choose mean queue size as our measure for evaluating a given approximation, we are interested in the percentage error

$$\%\Delta L_q = \frac{\hat{L}_q - L_q}{L_q} \cdot 100, \tag{1}$$

*This work was supported under Contract N00014-67-A-0214, Task 0001, Project NR 347 020, Office of Naval Research.

where $\hat{L}_q$ is the approximation to the true value L_q. Since Little's formula gives $L_q = \lambda W_q$, the same percentage error is incurred in estimating W_q. Corresponding percentage errors would be smaller in estimating L and W since $W = W_q + 1/\mu$ and $L = L_q + \lambda/\mu$, so using L_q yields conservative estimates of how well the approximations perform.

For single channel queues, the cost model we consider is one of determining the optimal mean service rate which "balances" customer wait against the cost of providing service and is given as

$$\min_{\mu} E[C(\mu)] = C_s\mu + C_w L \qquad (\$/\text{unit time}),$$

where C_s is the cost per unit time of providing service and is assumed proportional to μ, and C_w is the cost per unit time of customer "downtime." Normalizing by dividing by C_w gives the following cost optimization problem:

$$\min_{\mu} E[C(\mu)] = C\mu + L \qquad (C = C_s/C_w). \tag{2}$$

The percentage error of interest is

$$\%\Delta E = \frac{E[C(\hat{\mu}_0)] - E[C(\mu_0)]}{E[C(\mu_0)]}, \tag{3}$$

where $\hat{\mu}_0$ is the optimal value of μ based on $\hat{L}$ while μ_0 is the true optimal value of μ.

In the multichannel case we consider the cost problem of "balancing" server idle time against customer idle time (the model used by Brigham), namely,

$$\min_{c} E[C(c)] = C_s(c - \lambda/\mu) + C_w L_q \qquad (\$/\text{unit time}), \tag{4}$$

where c is the number of servers, C_s is now the marginal cost of a server per unit time, $(c - \lambda/\mu)$ is the percentage of server idle time and C_w is the cost per unit time of a customer waiting for service. Equation (4) can be normalized, and to be consistent with Brigham, we divide by C_s to get

$$\min_{c} E[C(c)] = c - \frac{\lambda}{\mu} + C^{-1}L_q \qquad (C = C_s/C_w). \tag{5}$$

2. SINGLE-CHANNEL CASES—EXPECTED-VALUE MEASURES OF CONGESTION

In this section we observe the performance of the $M/M/1$ as an approximation to $M/G/1$, $GI/M/1$ and $GI/G/1$, with respect to the criteria $\%\Delta L_q$.

2.1. *M/M/1 versus M/G/1*

The Pollaczek-Khintchine (P–K) formula allows one to calculate L_q for any $M/G/1$ queue, knowing only the coefficient of variation of the service distribution (call v_B) and the traffic intensity ρ, and yields

$$L_q = \frac{\rho^2(1 + v_B^2)}{2(1 - \rho)} \qquad (\rho \equiv \lambda/\mu), \tag{6}$$

where λ and μ are the mean arrival and service rates, respectively. For exponential service time, $v_B = 1$ and the P–K formula reduces to the familiar

$$L_q = \frac{\rho^2}{1-\rho}.$$

It is a simple matter now to calculate the decimal equivalent of the percent difference between $\hat{L}_q$ and L_q as

$$\Delta L_q \equiv \frac{\hat{L}_q - L_q}{L_q}$$

$$= \frac{\rho^2/(1-\rho) - \rho^2(1+v_B^2)/2(1-\rho)}{\rho^2(1+v_B^2)/2(1-\rho)}$$

$$= \frac{1-v_B^2}{1+v_B^2}.$$

Thus we see that the percentage error in using $\hat{L}_q$ rather than L_q depends only on v_B, and is independent of ρ in this case. A variety of cases for different v_B's were studied and the resulting $\%\Delta L_q$'s computed.* These are graphed in Figure 1. Also shown is the percent error for average system size, $\%\Delta L$.

We see that $\%\Delta L_q$ is rather sensitive to v_B and that the magnitude of $\%\Delta L_q$ is approximately equal to the percentage deviation from one of v_B. For example, a v_B 10 percent from one (0.9 or 1.1) produces a $\%\Delta L_q$ of about 10 percent. The sensitivity is not as great for $\%\Delta L$ (as we apriori knew to be the case) but is still sizable. The sensitivity with respect to L_q is greater for $v_B < 1$ where $M/M/1$ overestimates $M/G/1$ than for $v_B > 1$, where $M/M/1$ underestimates $M/G/1$. With respect to L, the sensitivity appears symmetrical. For example, a v_B of 0.5 yields $\%\Delta L_q \cong 60$ percent and $\%\Delta L = 35$ percent while a v_B of 1.5 yields 40 and 35 percent, respectively.

2.2. *M/M/1* versus *GI/M/1*

For a discrete k point distribution of interarrival times, it is not too difficult to compute L_q analytically for $GI/M/1$ (see chap. 5 of [2]). Table 1 gives 10 discrete distributions used in the analysis. All had the same mean ($1/\lambda = 2.9$), but different coefficients of variation (now denoted by v_A). Also given are the skewness (s) and kurtosis (k) values.†

These distributions were generated in the following manner. For v_A's less than 1, Erlang distributions were used and then discretized into k points (k varies betweea 3 and 5), with the value of the last point set to yield a mean of 2.9. Hence, the resulting v_A's differ somewhat from those of the original Erlangs due to both discretizing the distribution and placement of the final point to yield a common mean. For v_A's greater than 1, a hyper-exponential distribution was initially used and then discretized

*Sensitivity of L_q for $M/G/1$ is only of interest for the insight it may provide; the P–K formula is so easy to use it is not necessary to approximate L_q.

†Here, we use $s = \mu^3/\sigma^3$ and $k = \mu^4/\sigma^4$, where μ^n is the nth moment about the mean and σ is the standard deviation.

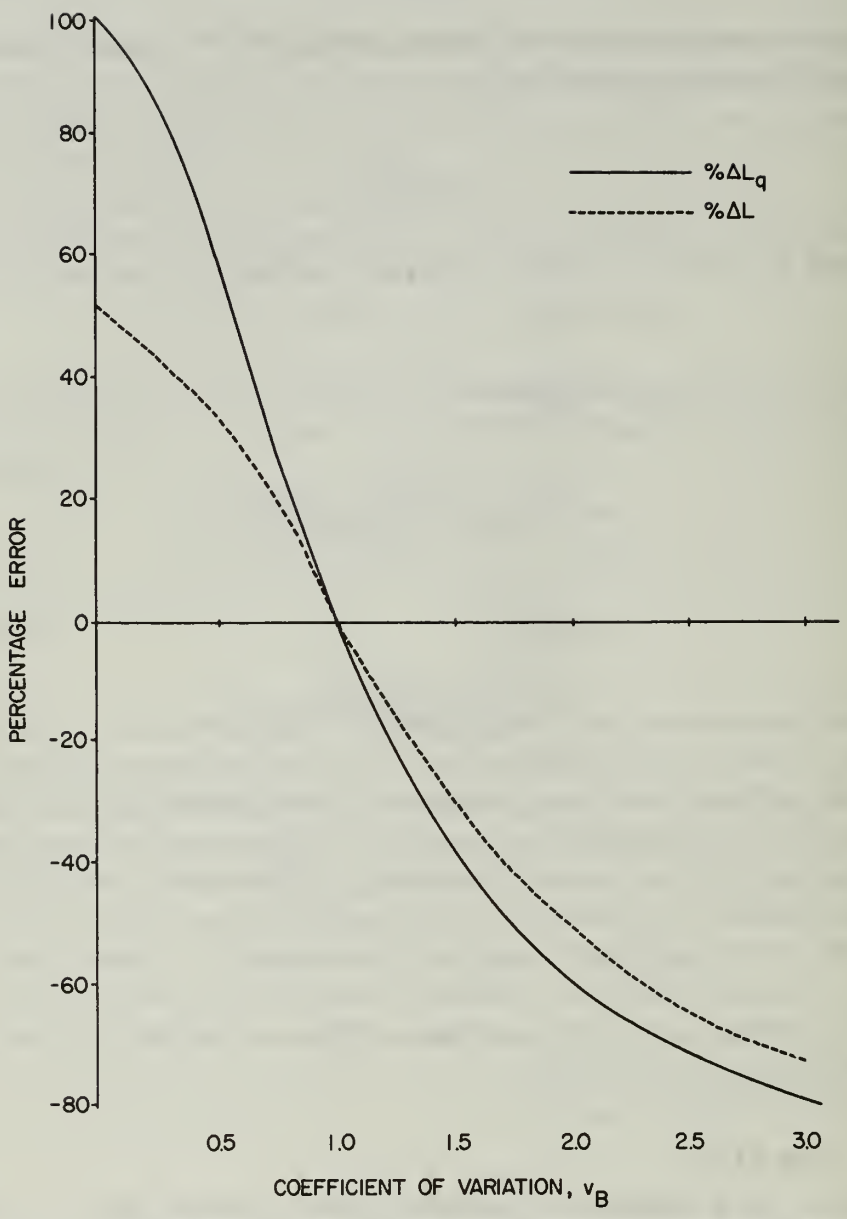


FIGURE 1. Percentage error versus service time coefficient of variation for $M/G/1$ cases.

in a manner similar to the Erlangs. By working with Erlangs and hyper-exponentials, it was possible to tell beforehand approximately what v_A's would result.

To obtain analytical results for $GI/M/1$ even when GI is a 3, 4, or 5-point discrete distribution is not particularly easy, but it can be done. It is necessary to first solve for the root of the generating function equation

$$\sum_{i=1}^{k} e^{-\mu t_i(1-z)} Pr\{T=t_i\} = z. \tag{8}$$

TABLE 1. *Discrete-point Distributions*

Mean = 2.0

No.	Moments[a]	Value	Probability	No.	Moments[a]	Value	Probability
1	$v_A = 0.28$ $s = -1.24$ $k = 4.30$	1 3 4	0.13 0.71 0.16	6	$v_A = 0.80$ $s = 1.36$ $k = 4.19$	1 3 5 7 10	0.50 0.245 0.135 0.07 0.05
2	$v_A = 0.40$ $s = -0.00$ $k = 3.03$	1 3 5	0.19 0.67 0.14	7	$v_A = 0.92$ $s = 1.30$ $k = 3.37$	1 3 5 7 9	0.55 0.21 0.09 0.04 0.11
3	$v_A = 0.47$ $s = 0.06$ $k = 2.14$	1 3 5	0.26 0.53 0.21	8	$v_A = 1.08$ $s = 2.03$ $k = 6.33$	1 3 5 6 12	0.58 0.21 0.09 0.04 0.08
4	$v_A = 0.55$ $s = 0.09$ $k = 1.60$	1 3 5	0.34 0.37 0.29	9	$v_A = 1.49$ $s = 3.08$ $k = 11.62$	1 3 5 7 19	0.65 0.20 0.06 0.03 0.06
5	$v_A = 0.61$ $s = 0.17$ $k = 1.41$	1 3 5 5.25	0.40 0.27 0.17 0.16	10	$v_A = 2.03$ $s = 4.33$ $k = 20.76$	1 3 5 7 31	0.70 0.19 0.05 0.02 0.04

[a] v_A = Coefficient of variation, s = measure of skewness, and k = measure of kurtosis.

It can be shown that there is only one root in (0, 1) and this is the desired root. Calling this root r_0, we can get the average queue size at arrival points (denoted by L_q^A) as

$$L_q^A = \frac{r_0^2}{1 - r_0}.$$

By semi-Markov process theory (see chap. 6 of [2]) one can show that the unconditional (general time) average queue size L_q is

$$L_q = \frac{\rho}{r_0} L_q^A = \frac{\rho r_0}{1 - r_0}. \tag{9}$$

Thus ΔL_q can be calculated from Equation (1) once r_0 is found. These computations were carried out utilizing a Newton-Raphson root-finding technique for various values of v_A and ρ (unlike $M/G/1$, ΔL_q here is not independent of ρ). Figure 2 shows the results for three settings of ρ and compares them with

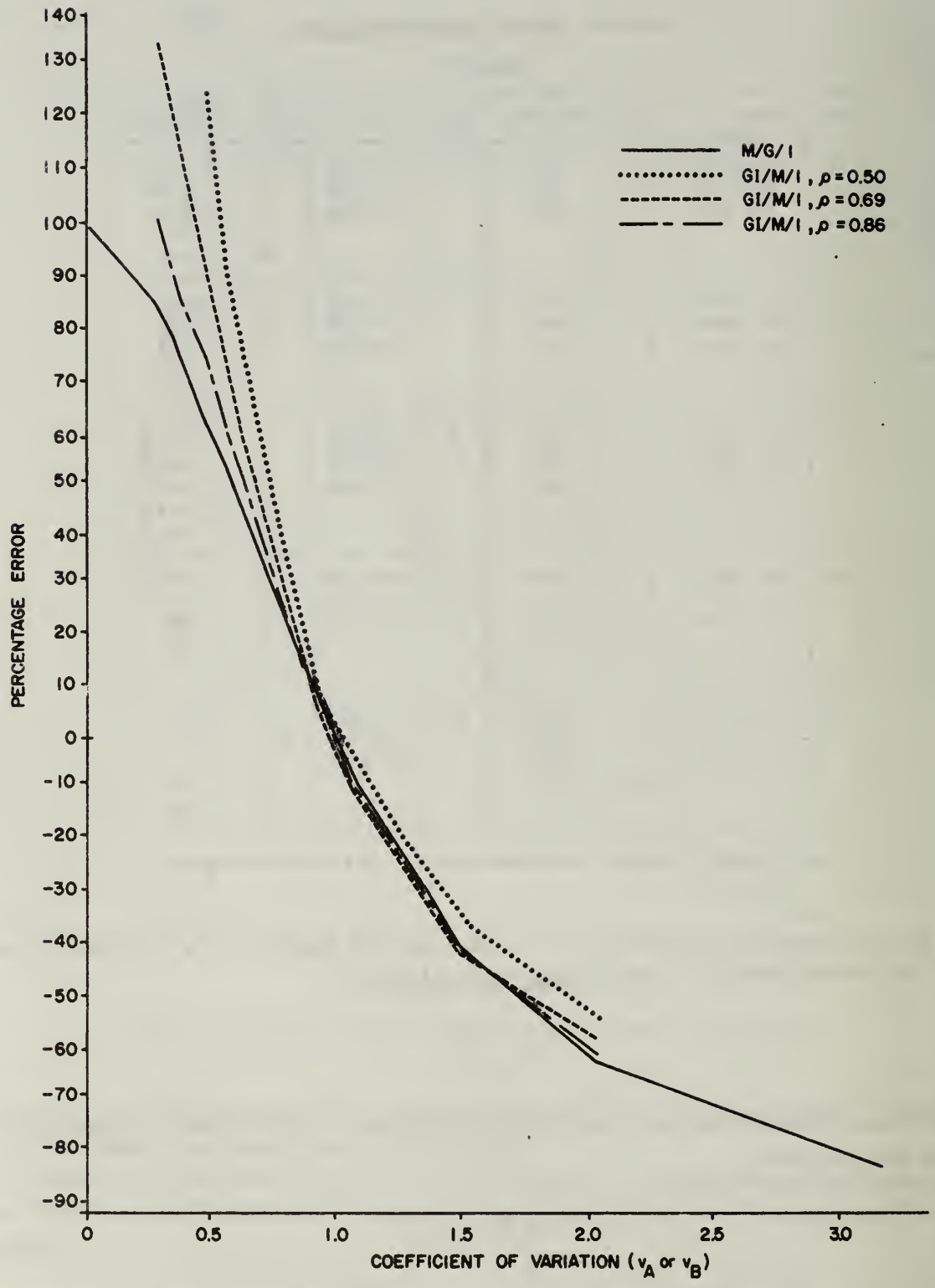


FIGURE 2. Percentage error in L_q versus coefficient of variation for $M/G/1$ and $GI/M/1$ cases.

the previous results for $M/G/1$. We see that ρ has only a slight effect and that the accuracy of the approximation for $GI/M/1$ is about the same as $M/G/1$.

2.3. $M/G/1$ versus $GI/G/1$

A single-channel queueing simulator was developed and utilized for this portion of the study. It was validated with pilot runs for $M/M/1$. Further, in these initial tests, run length and number of replications were varied to observe the resulting confidence interval estimates on L_q. The largest ρ to be studied was used in the pilot runs, so that conservative estimates of confidence and precision would result, since the larger the ρ, the greater the variation in L_q. Because $M/M/1$ was used for this analysis, when running the simulator with distributions having coefficients of variation greater than one, smaller confidence or larger intervals may result. However, this portion of the analysis was used only in deciding on run length and number of replications and in all $GI/G/1$ runs, confidence intervals were always calculated and are presented along with the mean value for L_q. The confidence intervals were always less than ± 10 percent around the mean L_q, and in *most* cases, less than ± 5 percent.

Table 2 shows the results for the initial pilot run. To balance running cost against precision, a run length of 10,000 customers with 10 replications was decided upon. Further, an initialization period of 500 customers was used to remove transient effects. One final pilot check was run for $GI/M/1$ with the results compared to the analytical program developed from the methodology explained in section 2.2 above. The case run was for $\rho = 0.69$, $v_A = 0.61$. The analytical procedure gave L_q as 0.958. The simulator gave a mean value of 0.941 with a 95 percent confidence interval of (0.901, 0.981), thus bracketing nicely the analytical result.

TABLE 2. M/M/1, $\rho = 0.86$; *Theoretical* $L_q = 5.283$

Run length (Number of customers)	Number of replications[a]	Mean L_q	95 percent confidence interval estimate	C.P.U. time (min)
500 (no initialization)	10	5.367	3.872–6.862 (±25%)	0.54
15,000 (no initialization)	10	5.233	4.835–5.630 (± 8%)	3.95
5,000 (1,000 initialization)	15[a]	5.117	4.683–5.551 (± 8%)	2.18
7,500 (500 initialization)	9[a]	5.151	4.741, 5.560 (± 8%)	2.04
7,500 (1,000 initialization)	10[a]	5.116	4.756, 5.476 (± 7%)	2.09
3,000 (500 initialization)	25[a]	5.116	4.744, 5.407 (± 7%)	2.09
10,000 (500 initialization)	8[a]	5.065	4.802, 5.324 (± 6%)	2.10

[a] Blocking (one long run divided into blocks) was actually used to avoid requiring initialization in every replication. The autocorrelation for blocks over 1,000 customers is negligible.

Three distributions were chosen from Table 1 for use in the simulator. Each was used for interarrival and service times in all possible combinations so that for a given ρ, there were nine cases. To achieve any particular ρ desired, the time values of the distributions chosen for service times were multiplied by ρ. Three cases of ρ were studied yielding a total of 27 cases run. The three distributions used were (from Table 1) numbers 1, 5, and 9, which provided a broad range of coefficients of variation. Additional distributions were also selected from Table 1 for study, but investigated at only one value of ρ. Results of the simulation runs are given in Table 3 (not all cases for $\rho = 0.69$ are shown).

TABLE 3. GI/G/1 *Mean* L_q *(95 percent lower and upper confidence limits)*

Case	v_A	v_B	$\rho=0.50$ ($\hat{L}_q=0.50$)	$\rho=0.69$ ($\hat{L}_q=1.55$)	$\rho=0.86$ ($\hat{L}_q=5.28$)
1×1	0.28	0.28	0.027 (0.026, 0.027)	0.074 (0.072, 0.077)	0.324 (0.313, 0.334)
1×3	0.28	0.47		0.137 (0.134, 0.140)	
1×5	0.28	0.61	0.041 (0.040, 0.042)	0.204 (0.198, 0.210)	1.020 (0.967, 1.073)
1×9	0.28	1.49	0.474 (0.456, 0.492)	1.559 (1.485, 1.632)	6.167 (5.359, 6.974)
1×10	0.28	2.03		3.014 (2.832, 3.196)	
3×1	0.47	0.28		0.191 (0.187, 0.195)	
3×3	0.47	0.47		0.269 (0.263, 0.275)	
3×5	0.47	0.61		0.349 (0.339, 0.359)	
3×9	0.47	1.49		1.742 (1.643, 1.841)	
3×10	0.47	2.03		3.069 (2.829, 3.309)	
5×1	0.61	0.28	0.119 (0.116, 0.121)	0.358 (0.351. 0.365)	1.187 (1.134, 1.241)
5×3	0.61	0.47		0.441 (0.429, 0.453)	
5×5	0.61	0.61	0.164 (0.160, 0.168)	0.520 (0.502, 0.537)	1.905 (1.815, 1.995)
5×9	0.61	1.49	0.585 (0.568, 0.601)	1.853 (1.772, 1.933)	6.763 (6.136, 7.390)
5×10	0.61	2.03		3.294 (3.042, 3.545)	
9×1	1.49	0.28	0.381 (0.370, 0.393)	1.712 (1.661, 1.763)	5.820 (5.468, 6.173)
9×3	1.49	0.47		1.856 (1.784, 1.928)	
9×5	1.49	0.61	0.515 (0.496, 0.534)	1.988 (1.905, 2.067)	6.455 (6.006, 6.906)
9×9	1.49	1.49	1.087 (1.066, 1.108)	3.356 (3.193, 3.519)	11.749 (10.648, 12.850)
9×10	1.49	2.03		4.505 (4.044, 4.967)	
10×1	2.03	0.28		3.075 (2.903, 3.247)	
10×3	2.03	0.47		3.219 (3.031, 3.407)	
10×5	2.03	0.61		3.337 (3.167, 3.506)	
10×9	2.03	1.49		5.314 (5.039, 5.590)	
10×10	2.03	2.03		6.680 (6.115, 7.246)	

Table 4 shows the resulting percentage errors for all runs for the three cases of ρ studied. These are mean percentage errors since they are based on the mean L_q from the simulation output.

To determine how much $\%\Delta L_q$ could vary, one could calculate 95 percent lower and upper confidence limits by using the 95 percent lower and upper confidence values on L_q in the $\%\Delta L_q$ calculation. To avoid having to do this we can relate the width of the confidence band on L_q to one on $\%\Delta L_q$ as follows.

Assume that β represents the width (± percent from mean) of the confidence interval on L_q. For example, from Table 3, the case of 9×9, $\rho = 0.86$ would have a β of 9.4 percent. Thus denoting by $L_{q,\,l}$ and $L_{q,\,\mu}$ as the lower and upper values in the confidence band on L_q, we have

$$L_{q,\,l} = \left(1 - \frac{\beta}{100}\right) L_q$$

and

$$L_{q,\,\mu} = \left(1 + \frac{\beta}{100}\right) L_q,$$

where L_q is the mean simulator value. Letting $\alpha_1 = (1 - \beta/100)$ and $\alpha_2 = (1 + \beta/100)$ the percentage errors based on $L_{q,\,l}$ and $L_{q,\,\mu}$ would be

TABLE 4. GI/G/1 $\%\Delta L_q$ *versus* v_A *and* v_B, $\rho = 0.69, 0.50$ *and* 0.86

v_A \ v_B	0.28	0.47	0.61	0.85	0.92	1.08	1.49	2.03
0.28	1950	1016	648				− 2	− 49
0.47	699	468	338				− 12	− 50
0.61	327	246	193				− 18	− 54
0.85				42	34	7		
0.92	97				15	4		− 59
1.08	57			4	4	− 14		− 62
1.49	− 10	− 18	− 23				− 55	− 66
2.03	− 50	− 53	− 54				− 71	− 77

v_A \ v_B	$\rho = 0.50$		
	0.28	0.61	1.49
0.28	1764	1121	5
0.61	322	204	− 15
1.49	32	− 3	− 54

v_A \ v_B	$\rho = 0.86$		
	0.28	0.61	1.49
0.28	1563	428	− 13
0.61	353	182	− 20
1.49	− 9	− 18	− 54

$$\%\Delta L_{q,\mu} = \frac{\hat{L}_q - L_{q,l}}{L_{q,l}} = \frac{\hat{L}_q - \alpha_1 L_q}{\alpha_1 L_q}$$

and

$$\%\Delta L_{q,l} = \frac{\hat{L}_q + \alpha_2 L_q}{\alpha_2 L_q}.$$

We desire the factors by which we must multiply $\%\Delta L_q$ to get $\%\Delta L_{q,l}$ and $\%\Delta L_{q,\mu}$, that is

$$\%\Delta L_{q,l} = f_l \%\Delta L_q$$

$$\%\Delta L_{q,\mu} = f_\mu \%\Delta L_q.$$

Thus,

$$f_l = \frac{\%\Delta L_{q,l}}{\%\Delta L_q} = \frac{(\hat{L}_q - \alpha_2 L_q)/\alpha_2 L_q}{(\hat{L}_q - L_q)/L_q} = \frac{\hat{L}_q/L_q - \alpha_2}{\alpha_2(\hat{L}_q/L_q - 1)},$$

but

$$\%\Delta L_q = \frac{\hat{L}_q - L_q}{L_q} 100$$

so that

$$\frac{\hat{L}_q}{L_q} = 1 + \frac{\%\Delta L_q}{100},$$

and therefore

$$f_l = \frac{100(1-\alpha_2)}{\alpha_2 \%\Delta L_q} + \frac{1}{\alpha_2}.$$

Likewise

$$f_\mu = \frac{100(1-\alpha_1)}{\alpha_1 \%\Delta L_q} + \frac{1}{\alpha_1}.$$

TABLE 5. *95-percent Confidence Range on* $\%\Delta L_q$, GI/G/1

Mean $\%\Delta L_q$	Confidence band width on L_q (β)					
	± 10 percent		± 8 percent		± 5 percent	
	$\%\Delta L_{q,l}$	$\%\Delta L_{q,\mu}$	$\%\Delta L_{q,l}$	$\%\Delta L_{q,\mu}$	$\%\Delta L_{q,l}$	$\%\Delta L_{q,\mu}$
−50	−55	−44	−54	−46	−52	−47
−25	−32	−17	−31	−18	−29	−21
−10	−18	0	−17	−2	−14	−5
−5	−13	6	−12	3	−10	0
−1	−10	10	−8	8	−6	4
1	−8	12	−6	10	−4	6
5	−5	17	−2	14	0	11
10	0	22	2	20	5	16
25	14	39	16	36	19	32
50	36	67	39	63	43	58

Table 5 gives, for various $\%\Delta L_q$, $\%\Delta L_{q,l}$ and $\%\Delta L_{q,\mu}$ for $\beta = 10$, 8, and 5 percent (β was always less than 10 percent for all cases). Thus looking at Table 5, realistically we should not consider $\%\Delta L_q$ values in Table 4 any "finer" than ±5 percent (or perhaps even ±10 percent) increments.

Looking once more at Table 4, we see that ρ has little effect, and that for v_A and v_B both small we get large positive errors ($\hat{L}_q >> L_q$), while for v_A and v_B both large we get large negative errors. This appears consistent with previous *M/G/*1 and *GI/M/*1 results indicating $\hat{L}_q$ overestimates for v_A and $v_B < 1$ and underestimates for v_A and $v_B > 1$. The smallest errors appear to occur when a large v_A is balanced by a small v_B or a small v_A balanced by a large v_B or both v_A and v_B near one. In assessing the sensitivity of *M/M/*1 for *GI/G/*1 situations, the simplicity in comparing the percentage error in L_q to the degree of deviation from one of the nonexponential coefficient of variation vanishes since both coefficients of variation differ from one. However, we would expect small percentage errors when both v_A and v_B are near one so that some measure, such as $v_A v_B = 1$ or $v_A + v_B = 2$, is a natural extension. Figure 3 plots absolute percentage error versus $v_A + v_B$ and we can see the general sensitivity to $v_A + v_B$. For any reasonable percentage error (say $\leq \pm 25$) $v_A + v_B$ should lie between 1.7 and 2.2.

Implicit in the discussion thus far is that the error in L_q is due solely to the coefficient of variation; that is, there is no effect due to higher moments. This is certainly true for *M/G/*1 as can be seen from Equation 7. However, there may be some effect of higher moments in the other cases (*GI/M/*1 and/or *GI/G/*1). To investigate this possibility, three special *GI/M/*1 cases were studied and the results are presented in Table 6. The first special case involved a two-point distribution with $v_A = 1.0007$ (very close to 1). Even though v_A is almost identical to that of the exponential, using only two points causes the distribution itself to be fairly nonexponential (a great difference in the higher moments). The other two distributions studied were a 20-point and a 60-point approximation of an exponential. The 20-point distribution had a resulting $v_A = 0.970$ (but was much closer to exponential than its two-point counterpart with $v_A = 1.0007$), while the 60-point distribution had a resulting $v_A = 0.995$. From Table 6, it appears there is only a slight effect due to higher moments except when ρ is very small ($\rho = 0.20$) as the percentage errors of the two-point distribution are not appreciably larger than the 60- and 20-point approximations.

3. SINGLE-CHANNEL CASES, COST OPTIMIZATION

The single-channel cost model of Equation (2) was investigated when using *M/M/*1 as an approximation for *M/G/*1 and *GI/M/*1, respectively.

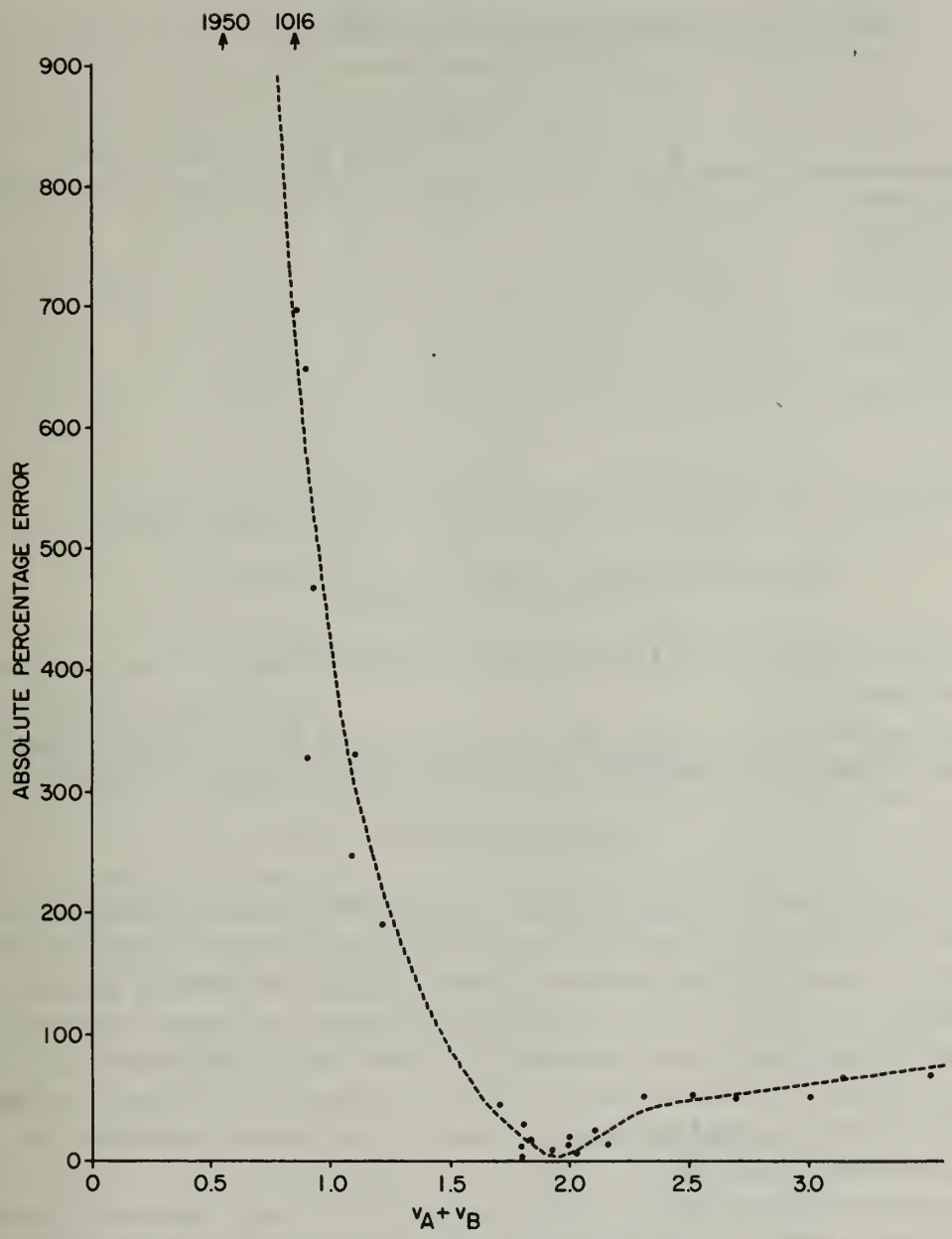


FIGURE 3. Absolute percentage error in L_q versus $v_A + v_B$, p = 0.69 for $GI/M/1$ cases.

For $M/G/1$, to determine $\%\Delta E$, the cost error, it is first necessary to find the optimal service rate μ_0. Employing Equation (6) in Equation (2) gives

$$E[C(\mu)] = C\mu + \frac{\rho^2 + (1 + v_B^2)}{2(1 - \rho)} + \rho \tag{10}$$

$$= C\mu + \frac{2\lambda\mu + \lambda^2(v_B^2 - 1)}{2\mu(\mu - \lambda)}$$

TABLE 6. *Effect of Higher Moment Study*, GI/M/1 $\%\Delta L_q$ *Versus* v_A *and* ρ

v_A	s	k	ρ					
			0.20	0.34	0.50	0.69	0.86	0.96
1.000 (True Expon)	2	9	0	0	0	0	0	0
0.995 (60-point Expon)	1.98	8.72	4	2	1	1	1	5
0.970 (20-point Expon)	1.84	7.23	20	8	5	4	4	9
1.001 (2-point Distbn)	1.15	2.33	90	6	−6	−5	−1	7

Taking $dE/d\mu$ and setting it equal to zero yields the following quartic equation in μ:

$$4C\mu^4 - 8\lambda C\mu^3 + 4\lambda(\lambda C - 1)\mu^2 + 4\lambda^2(1 - v_B^2)\mu + 2\lambda^3(v_B^2 - 1) = 0. \tag{11}$$

Thus we must find the four roots and choose the one which satisfies $\mu > \lambda$ and minimizes the right-hand side of Equation (10).

We can now see an advantage in using $M/M/1$ rather than $M/G/1$. If G can be approximated by M, Equation (11) reduces to a quadratic and the optimal μ is found by the neat expression

$$\hat{\mu}_0 = \lambda + \sqrt{\lambda/C}. \tag{12}$$

For a variety of cases, we have actually solved for the four roots of Equation (11) and thus determined μ_0. For these same cases, $\hat{\mu}_0$ was computed from Equation (12) and the resulting $\%\Delta E$ calculated from Equation (3) using Equation (10), first with $\mu = \hat{\mu}_0$ and then with $\mu = \mu_0$.

With respect to $\%\Delta E$ for $GI/M/1$, to find the true optimal μ requires a numerical procedure. For a particular value of μ, r_0 can be computed, then L_q, and finally, using Equation (2), $E[C(\mu)]$ can be obtained. To find the μ which minimizes $E[C(\mu)]$, μ was stepped off in 0.01 increments starting with $\mu = \lambda + 0.01$, $E[C(\mu)]$ computed for each μ as described above, and the minimizing value of μ which yielded the smallest $E[C(\mu)]$ selected.

Figure 4 shows $\%\Delta E$ when using $\hat{L}_q$ instead of L for the $M/G/1$ cases. Also shown are the previously graphed quantities $|\%\Delta L_q|$ and $|\%\Delta L|$ when using $\hat{L}_q$. It is immediately obvious that there is an order of magnitude reduction in percentage error when the approximant is used in the cost optimization model. Twenty-three values of λC were tried (it turns out for the $M/G/1$ case that $\%\Delta E$ depends on the product λC and not on λ and C individually) from 0.01 to 100.00 and the *worst* case (largest $\%\Delta E$) is graphed.

Similar results were obtained for the $GI/M/1$ cases run, and thus they are not graphed. No cost optimization runs were made for $GI/G/1$ due to the expense of searching for the true optimal μ using the simulator, but it is reasonable to assume similar reductions in percentage errors would result. This phenomenon of relative insensitivity of cost functions near the optimum region is not new. It has long been recognized in inventory modeling. In light of this, these results are not surprising, only comforting.

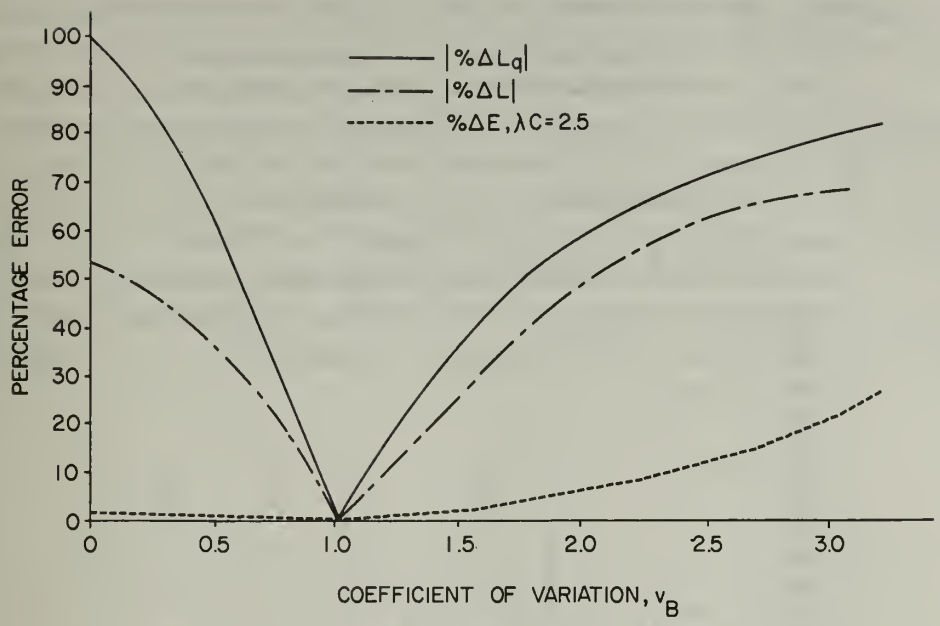


FIGURE 4. Percentage errors in L, L_q and costs, $M/G/1$ cases.

4. MULTICHANNEL RESULTS

We now consider using an $M/M/c$ model as an approximation for certain $GI/G/c$ situations. A c-channel simulator was developed and validated in a manner similar to the single-channel simulator described in section 2.3. Unfortunately the arbitrary number of servers introduces a whole new dimension to the analysis and increases the running time by an order of magnitude. Thus the results reported here are less extensive than those for the single-channel cases, but nevertheless should provide insight when taken together with their single-server counterparts.

It was supposed in the following experiments that the service time distribution was always the empirical service time distribution as observed by Brigham, which had a coefficient of variation of .29. Two sets of simulations were then performed. First, an exponential interarrival distribution was assumed. Hence the interarrival coefficient of variation is unity (i.e., $M/G/c$). The traffic intensity ($\rho = \lambda/c\mu$) was varied from a low of 0.2 to a high of 0.8. Several different cases with two, three, or four servers were studied and the corresponding percent errors were calculated. Table 7 (left portion) contains the results from this initial collection of experiments. It can be seen that the simplifying exponential assumption in these cases causes the mean wait in queue to be underestimated by a factor of about one-fourth. There appears to be no dependency of the errors on ρ or c. So one might reasonably expect this fractional underestimate to hold for an increasing number of servers. It is also remarkable that the size of the error is not excessive in view of the fact that the service-time distribution is significantly different from the exponential by a chi-squared test. However, this is rather consistent with the $M/G/1$ results, for Figure 1 shows approximately a 25 percent error for a $v_B = 1.3$.

Table 7 (right portion) provides comparable data from the second collection of experiments. In this set a hyperexponential distribution with a coefficient of variation equal to 1.14 was used as the interarrival distribution. This particular distribution was selected because it matched closely the empirical arrival data collected by Brigham. The mean interarrival time and the number of servers were varied precisely as in the first collection of simulations.

TABLE 7. *Percentage Error When* M/M/c *Is Used to Approximate* L_q

ρ	For *M/G/c* with the use of Brigham's data $v_B = 1.29$			For *GI/G/c* with the use of Brigham's data $v_A = 1.14$, $v_B = 1.29$		
	Number of servers (*c*)			Number of servers (*c*)		
	2	3	4	2	3	4
0.20	−21			−38		
0.33	−19	−24			−31	
0.38			−15			
0.39	−29					
0.40			−32			−34
0.42		−24			−30	
0.45	−26		−17			−35
0.50		−21	−25		−31	−38
0.55			−15			−32
0.58		−21			−34	
0.60			−24			−29
0.63	−24					
0.65			−16			−35
0.67		−28			−37	
0.71	−26					
0.75		−20				
0.80	−32					

The errors in this second situation show that the approximation has now deteriorated to the point where it underestimates the mean wait in queue by a factor of about one-third. However, this again is consistent with the behavior we observed in the single-server cases shown in Figure 3 for corresponding v_A and v_B. The approximation appears to be independent of the traffic intensity, to underestimate for high coefficients of variation and to overestimate for low coefficients of variation. In order to reinforce this latter conjecture, two of the basic distributions from Table 1, namely, distributions 1 (coefficient of variation = 0.28) and 10 (coefficient of variation = 2.03), were selected for further Monte Carlo experiments parallel to the above. These distributions have either extremely small or extremely large coefficients of variation. So they should exhibit the "over-underestimate" behavior.

TABLE 8. *Percentage Error When Using* M/M/c *to Approximate* L_q *For Some Discrete* GI/G/c *Cases*

$\rho = 0.69$

v_A \ v_B	$c = 1$		$c = 2$		$c = 3$		$c = 4$	
	0.28	2.03	0.28	2.03	0.28	2.03	0.28	2.03
0.28	1,950	−49	3,023	−42	3,685	−36	4,289	−44
2.03	−50	−77	−57	−75	−56	−126	−67	−84

Table 8 contains these results and those of the comparable single-server cases from Table 4 and the expected does occur. Hence we conclude that the *M/M/c* as an approximation to *GI/G/c* may be

cceptable provided the interarrival and service-time coefficients of variation are not too far from one. he magnitude of the error involved depends on the difference between the coefficients of variation and nity (say the sum of v_A and v_B versus two) and is essentially independent of ρ and c.

The multiserver cost optimization model given by Equation (4) was studied using $M/M/c$ to approxiate L_q for some $GI/G/c$ cases involving Brigham's data ($v_A = 1.14$, $v_B = 1.29$). Figure 5 shows the esults. The solid curves represent the cost, $E[C(c)]$ for different values of c as $r = \lambda/\mu$ is varied. he ratio $C = C_s/C_w$ is fixed at two. Changing C gives curves of similar shape but displaced up or own, thus we have presented results for one setting of C only.

Since we desire the minimum expected cost, the curves indicate the number of servers for a given The optimum cost curve is "scalloped," so for any particular r, we project up to the "scallop" to

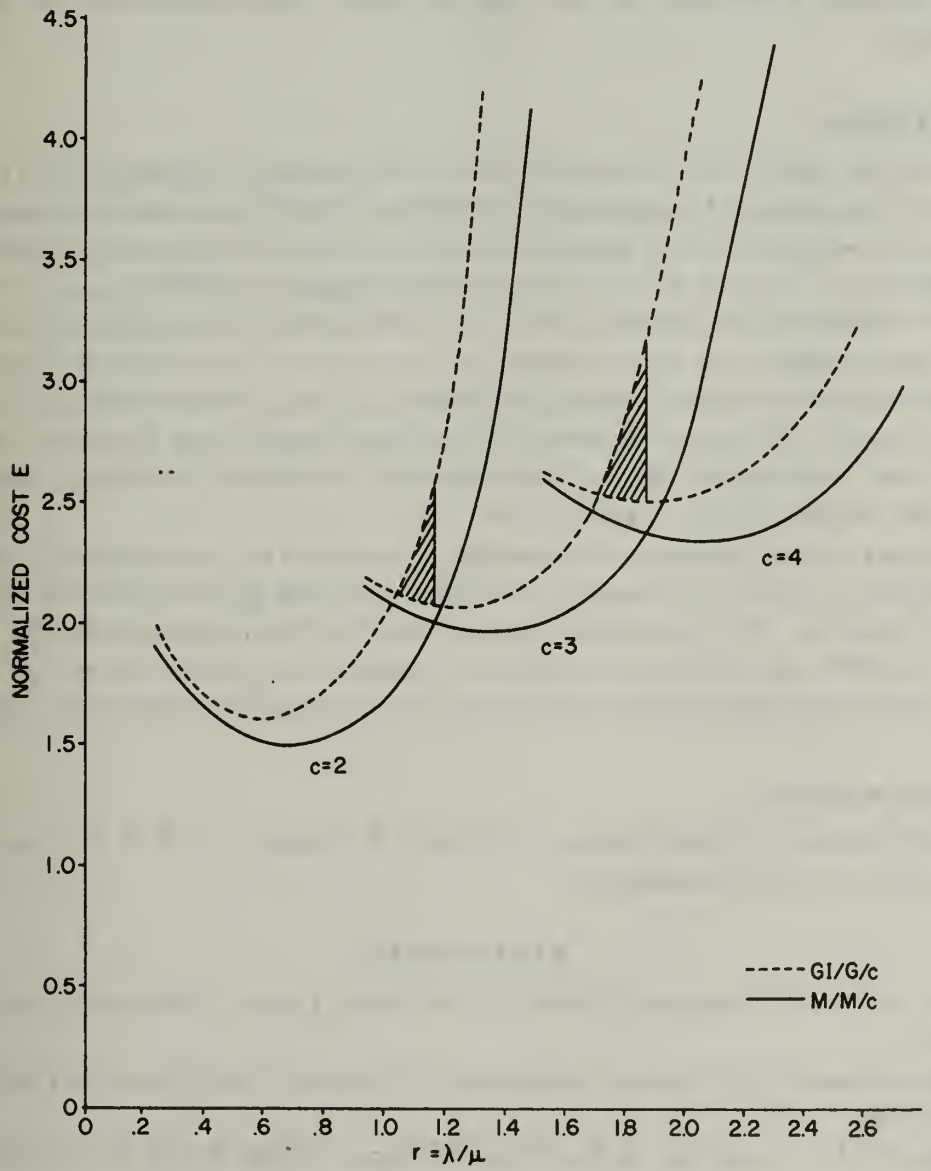


FIGURE 5. Percentage errors in costs for a $GI/G/c$ case.

find the optimum number of servers. For example, for $r \leq 1.18$, the optimal c by using $M/M/c$ is two, from $1.18 < r \leq 1.87$, the optimal c is three, etc.

For the true $GI/G/c$ case, the dashed curves give the optimal number of servers and the true expected cost. Thus we see for r less than 1.03, the $M/M/c$ gives the correct optimal number of servers but an underestimate of the actual expected costs. However, for r between 1.03 and 1.18, the $M/M/c$ is in error on the optimal number of servers, since the true optimal number of servers jumps to three. The cross-hatched portions show the actual error in the true expected costs for using the incorrect number of servers. The maximum percentage error appears to be approximately 25 percent, at $r = 1.18$. The same situation holds true for r between 1.70 and 1.83, with a maximum percentage error of about 25 percent for $r = 1.87$.

Over most values of r, $M/M/c$ gives correct results with respect to the optimal number of servers. For those cases when $M/M/c$ is in error, the $\%\Delta E$ is at most 25 percent and is only that high for very small ranges of r.

5. CONCLUSIONS

This study has shown that for estimating mean value measures of congestion, there is a sizeable sensitivity to the assumption of exponentiality. For $M/G/1$ and $GI/M/1$ cases, the percentage error when $M/M/1$ is used to estimate the mean queue wait is roughly equal to the percentage deviation from one of the coefficients of variation of the nonexponential distribution. In $GI/G/1$ cases, the percentage error seems to depend on the deviation of the sum of the coefficients of variation from two, with the percentage error roughly twice the percentage deviation from two. These "rules of thumb" are only valid for percentage errors up to 50 percent (see Figures 1, 2 and 3). Similar results appear to hold for multichannel models. The number of servers and the traffic intensity have little effect on the sensitivity. The same is true for the effect of higher moments, so that the coefficient of variation of the nonexponential distribution(s) is the key indicator.

With respect to cost optimization, the sensitivity of the results to the assumption of exponentiality is greatly diminished. Thus even though one has sizable percentage errors in estimation of mean value measures of congestion, the degradation in optimal costs in a cost optimization situation when M/M is used in non-M/M cases is smaller by an order of magnitude. This reflects the not uncommon phenomenon of the relative flatness of the optimal cost curve in a broad region about the optimal.

ACKNOWLEDGMENT

The author wishes to express his appreciation to K. L. Chhabra and H. D. Kahn for their aid in computer programming and computation.

REFERENCES

[1] Brigham, G., "On A Congestion Problem In An Aircraft Factory," Operations Res. *3*, 412–428 (1955).

[2] Gross, Donald and Carl M. Harris, *Fundamentals Of Queueing Theory* (John Wiley and Sons, New York, 1974).

[3] Kingman, J. F. C., "Inequalities In The Theory Of Queues," J. Roy. Statist. Soc. Ser. B *32*, 102–110 (1970).

[4] Marshal, Kneale T., "Some Inequalities In Queueing," Operations Res. *16*, 651–665 (1968).

[5] Marshal, Kneale T., "Bounds For Some Generalizations Of The *GI/G*/1 Queue, Operations Res. *16*, 841–848 (1968).

[6] Morse, Phillip M., *Queues, Inventory And Maintenance* (John Wiley and Sons, New York, 1958).

INFERENCE FROM ACCELERATED LIFE TESTS USING EYRING TYPE RE-PARAMETERIZATIONS*

Nozer D. Singpurwalla, Victor C. Castellino, and Donald Y. Goldschen

The George Washington University
School of Engineering and Applied Science

ABSTRACT

Least squares estimators of the parameters of the generalized Eyring Model are obtained by using data from censored life tests conducted at several accelerated environments. These estimators are obtained after establishing that the Gauss-Markov conditions for least squares estimation are satisfied. Confidence intervals for the hazard rate at use conditions are obtained after empirically showing that the logarithm of the estimate of the hazard rate at use conditions is approximately normally distributed. The coverage probabilities of the confidence intervals are also verified by a Monte Carlo experiment. The techniques are illustrated by an application to some real data.

1. INTRODUCTION

In certain life testing experiments it is a common practice to subject the test items to conditions of high stress. Such tests are called *accelerated life tests.* This situation prevails when an inordinate amount of time is required to obtain a failure at use conditions stress. In several instances, based upon a knowledge of the physics of failure of a device, it is possible to assume a relationship between the parameters of the failure distribution and the stresses (see Singpurwalla [8], [9], [10]). The Eyring Model is one such relationship which is frequently used in practice; for example, in the accelerated life testing of tantalum capacitors (see Thomas [11]).

A range of thermal and nonthermal stresses is prescribed and the exponential model for failure times for all values of the stress within this range is assumed. The exponential scale parameter λ, the hazard rate, is expressed as a function of a thermal stress T_i and a nonthermal stress V_i, as prescribed by the Eyring Model. The model is developed by using concepts of statistical mechanics (cf. Shooman [7], p. 454) which express the time rate of parameter degradation in terms of the given stress pair (T_i, V_i).

As ascertained by Pieruschka [6] we have assumed here that the severity of the stress level does not change the form of the lifetime distribution, but that the stress levels have an influence on the values of the parameters. Based upon the results of a life test conducted at $m \geq 4$ distinct accelerated values of the stresses $\{T_i, V_i\}$, $i = 1, 2, \ldots, m$ we would like to predict λ_u, the value of the hazard rate λ at use conditions stress (T_u, V_u).

---

*This research was sponsored in part by the Office of Naval Research under Contract N00014-67-A-0214, Task 0001, Project NR 347 020, and by the Air Force Aerospace Research Laboratories, Air Force Systems Command, United States Air Force, Contract F 33615-74-C-4040.

2. PRELIMINARIES

Suppose that the device under consideration is subjected to the constant application of a thermal stress T_i and a nonthermal stress V_i, and that its failure distribution is given by $f(t; \lambda_i) = \lambda_i \exp(-\lambda_i t)$, $\lambda_i > 0$, $t \geq 0$. A re-parameterization of λ_i as given by the Eyring Model is

$$\lambda_i = AT_i \exp(-B/KT_i) \exp(CV_i + DV_i/KT_i), \tag{2.1}$$

where K is Boltzman's constant having a value of 1.38×10^{-16} ergs/degree, T_i and V_i are the stresses, and A, B, C, and D are the unknown parameters. One needs to estimate these parameters and subsequently obtain point and interval estimates of λ_u, the hazard (failure) rate at some use conditions stresses (T_u, V_u).

Based on censored sample life tests conducted at $m \geq 4$ different stress levels $\{T_i, V_i\}$, $i = 1, 2, \ldots, m$, $i \neq u$, point estimators of A, B, C, and D are obtained using the method of least squares, where the m values of (T_i, V_i) are sufficiently high to induce failures. To insure that there is no correlation between the tests, it is desirable to use a randomization scheme. A reasonable procedure is to choose the m stress pairs (T_i, V_i) at which it is desired to run the life tests; then, to assign random numbers to the m stress pairs, and finally to run the life test according to this random sequence.

Let n_i items be put on the ith life test for which the stress levels are fixed at (T_i, V_i). If this test is terminated after r_i failures have occurred, then a minimum variance, unbiased, efficient and sufficient estimator of the mean time to failure θ_i (where $\theta_i = 1/\lambda_i$) is $\hat{\theta}_i$, given by Epstein and Sobel [2] as

$$\hat{\theta}_i = \left(\sum_{j=1}^{r_i} t_{ij} + (n_i - r_i) t_{ir_i} \right) / r_i,$$

if there is no replacement as items fail, and as

$$\hat{\theta}_i = (n_i t_{r_i}) / r_i,$$

if there is replacement as the items fail. Here the t_{ij}, $j = 1, 2, \ldots, r_i$, denote the observed r_i times of failure.

Also the probability density function of $\hat{\theta}_i$, $g(\hat{\theta}_i)$ is given as

$$g(\hat{\theta}_i) = \frac{\exp(-r_i \hat{\theta}_i / \theta_i)}{\Gamma(r_i)} (r_i/\theta_i)^{r_i} (\hat{\theta}_i)^{r_i - 1} \qquad \begin{cases} \hat{\theta}_i \geq 0 \\ r_i \geq 1 \end{cases}.$$

Let $\hat{\lambda}_i = 1/\hat{\theta}_i$ be an estimate of the hazard rate λ_i. Then, it can be shown (see Singpurwalla [8]), that the distribution of $\hat{\lambda}_i$ is given by

$$g(\hat{\lambda}_i) = \exp(-r_i/\theta_i \hat{\lambda}_i) (\Gamma(r_i))^{-1} (r_i/\theta_i)^{r_i} (\hat{\lambda}_i)^{-r_i - 1}, \quad \hat{\lambda}_i, r_i > 0$$

$$= 0, \text{ otherwise.}$$

This test is conducted for all the $m \geq 4$ values of (T_i, V_i), and based on the data $\{\hat{\lambda}_i, T_i, V_i, r_i, n_i\}$, $i=1, 2, \ldots, m$, it is desired to obtain an estimator of λ_u, the hazard rate at use conditions stress (T_u, V_u).

For the procedures discussed in section 3, we will have to consider the mean and the variance of $\hat{Y}_i = \log_e \hat{\lambda}_i$. The moment generating function of $\hat{Y}_i$, $M_{\hat{Y}_i}(t)$, can be obtained as

$$M_{\hat{Y}_i}(t) = \left(\frac{r_i}{\theta_i}\right)^t \frac{\Gamma(r_i - t)}{\Gamma(r_i)}.$$

Thus it follows that

(2.2) $$E(\hat{Y}_i) = \log_e\left(\frac{r_i}{\theta_i}\right) - \psi(r_i)$$

and

(2.3) $$\text{Var}\,(\hat{Y}_i) = \psi'(r_i),$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and the trigamma functions, respectively, tabulated in Abramowitz and Stegun [1] pp. 258–273.

We can easily verify that the exact distribution of $\hat{Y}_i$ is not normal. However, based on a Monte Carlo experiment, we can reasonably well assume that $\hat{Y}_i$ is approximately normally distributed. The results of this Monte Carlo experiment are discussed in the appendix.

3. LEAST SQUARES ESTIMATION FOR THE EYRING MODEL

In this paper we obtain the least squares estimators of the parameters of the Eyring Model. The maximum likelihood estimators require a solution of four simultaneous nonlinear equations, and this is not very desirable.

We first define the following quantities:*

$\vec{\alpha}$: $[\log A, C, D, -B]$,
$\vec{\epsilon}$: $[\log \epsilon_1, \log \epsilon_2, \ldots, \log \epsilon_m]$,
$\vec{Z}$: $[z_1, z_2, \ldots, z_m]$, where $z_i = \log(\hat{\lambda}_i / r_i T_i) + \psi(r_i)$, and

$$X = \begin{bmatrix} 1 & V_1 & V_1/KT_1 & (KT_1)^{-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & V_m & V_m/KT_m & (KT_m)^{-1} \end{bmatrix}$$

If we assume a multiplicative error term for our model given by Equation (2.1), then a model of observations can be written in matrix form as

$$\vec{Z} = \vec{X}\vec{\alpha} + \vec{\epsilon}.$$

*All logarithms are to the base e.

Since $E(\log \hat{\lambda}_i) = \log \lambda_i + \log r_i - \psi(r_i)$, $\mathrm{Var}(\log \hat{\lambda}_i) = \psi'(r_i)$, and because of randomization of the life test experiment

i) $E(\vec{\epsilon}) = 0$, and
ii) $E(\vec{\epsilon}\,\vec{\epsilon}') = \mathbf{W} = \mathrm{diag}[\psi'(r_1), \ldots, \psi'(r_m)]$,

where $\vec{\epsilon}'$ denotes the transpose of $\vec{\epsilon}$ and $\mathrm{diag}[\psi'(r_1), \ldots, \psi'(r_m)]$ is an $m \times m$ diagonal matrix. Of course $\mathbf{W}$ is the variance-covariance matrix of the ϵ_i, and since the $\psi'(r_i)$ values have been tabulated, the matrix $\mathbf{W}$ is known.

The above results allow us to invoke an application of the Gauss-Markov Theorem (Graybill [3], p. 115), and obtain the best linear unbiased estimators (BLUE) of $\vec{\alpha}$, $\vec{a}$. Furthermore, if the distribution of $\vec{Z}$ is a multivariate normal, confidence intervals for $\vec{\alpha}$ and λ_u can be easily obtained. Though an analytical result justifying the assumption of normality for $\vec{Z}$ was not possible, a Monte Carlo type investigation suggested the approximate normality of $\vec{Z}$. Under the assumption of the approximate normality of $\vec{Z}$, we obtain confidence limits for $\vec{\alpha}$ and λ_u based on the fact that $\vec{a}$ and $\hat{\lambda}_u$ (the least squares estimator of λ_u) are approximately normally distributed. We subsequently test the accuracy of our confidence intervals (and hence the validity of our assumption), by simulating several such experiments and computing the probability of coverage of our confidence limits. The results of this simulation are presented in the appendix and these do not refute our assumption.

Let a_i, $i = 1, 2, 3, 4$ be the least squares estimators of $\log A$, C, D, and $-B$, respectively, and let $\vec{a}$: $[a_1, a_2, a_3, a_4]$. Then the Gauss-Markov Theorem (Graybill [3], pp. 113–122) asserts the following results:

1. The minimum variance, unbiased, consistent and efficient linear estimator of the vector $\vec{\alpha}$ is given by $\vec{a}$, where

$$\vec{a} = (X'\mathbf{W}^{-1}X)^{-1}X'\mathbf{W}^{-1}\vec{Z}. \tag{3.1}$$

2. A best linear unbiased estimator (the predicted value) of λ_u, the hazard rate at use conditions stress (T_u, V_u) is $\bar{\lambda}_u$, where

$$\bar{\lambda}_u = T_u \exp\left[a_1 + a_2 V_u + \frac{a_3 V_u}{KT_u} + \frac{a_4}{KT_u} \right]. \tag{3.2}$$

Under the assumption that $\vec{Z}$ is approximately normally distributed, a $100(1-\alpha)$ percent prediction interval for λ_u is given by the following expression.

$$P[\bar{\lambda}_u \exp(-t^*_{\alpha/2} J_u) \leq \lambda_u \leq \bar{\lambda}_u \exp(t^*_{\alpha/2} J_u)] = 1-\alpha, \tag{3.3}$$

where

$$J_u = s\sqrt{1 + X_u'(X'\mathbf{W}^{-1}X)^{-1}X_u}; \qquad X_u: \left(1, V_u, \frac{V_u}{KT_u}, \frac{1}{KT_u}\right),$$

and s^2, an estimator of the residual variance, is given by

$$s^2 = \{\vec{Z}'P(I - XP(X'\mathbf{W}^{-1}X)^{-1}P'X')P'\vec{Z}\}/(m-4).$$

In the above expressions, $t^*_{\alpha/2}$ is the $100(\alpha/2)$ percentage point of a student's t distribution with $n-4$ degrees of freedom, I is the identity matrix, and P is such that $P'WP=I$.

4. AN ILLUSTRATIVE EXAMPLE

The methods of the previous sections will be illustrated here by an example. Table 1 contains the actual times to failure of a life test conducted on tantalum capacitors when they are subjected to a voltage stress and a temperature stress.

TABLE 1. *Tantalum Capacitor Failure Times Under Voltage And Temperature Stresses*

Test number (index i)	Number on test n_i	Voltage V_i	Temperature °C T_i	Actual times to failure (hr)
1	1,000	35.0V	85 °C	20, 90, 700, 37000
2	200	40.6V	85 °C	20, 3600, 9500, 27000
3	50	46.5V	85 °C	800, 2800
4	50	51.5V	85 °C	500, 800, 2400, 10700, 10700
5	500	46.5V	45 °C	100, 1200, 7500, 20000, 26000, 27300
6	175	46.5V	5 °C	1000
7	175	62.5V	5 °C	25, 50, 165, 500, 620, 720, 820, 910, 980, 1270, 1600, 2270, 2370, 4590, 4690, 4880, 7560, 8730, 12500
8	50	57.0V	45 °C	8900

We obtain the least squares estimator of $\vec{\alpha}$, $\vec{a}$ as

$$\vec{a}: [-41.76, \quad 0.529, \quad -0.00017\times 10^{-16}, \quad 0.00728\times 10^{-16}].$$

We next predict $\log \bar{\lambda}_i$, $i=1, 2, \ldots, 8$ by using the relationship

$$\bar{\lambda}_i = T_i \exp\left(a_i + a_2 V_i + \frac{a_3 V_i}{KT_i} + \frac{a_4}{KT_i}\right)$$

and compare these with the actual values of $\log \hat{\lambda}_i$ that went into the estimation of $\vec{\alpha}$.

We obtain the 95-percent prediction intervals for $\log \lambda_i$ by assuming normality of $\log \hat{\lambda}_i$, $i=1, 2, \ldots, 8$, and using an expression corresponding to Equation (3.3).

The results of our estimation are given in Table 2, from which it appears that the predictions compare favorably well with the actual values. All the same, in order for us to test whether the Generalized Eyring Model is indeed acceptable in this particular situation, we proceed as follows. Since tests 1 and 2 and tests 3 and 4 are proximately close, we use them to obtain an estimate of the variance of the error; this is 2.3437. We next compute the mean squared error due to lack of fit, and this is 1.4564. The ratio of these quantities, 1.6092, is compared to the 95-percent percentile point of an F-distribution with 2 and 3 degrees of freedom, respectively; this is 9.55. Hence we do not reject the hypothesis that the fit of the Generalized Eyring Model is adequate.

TABLE 2. *Comparison of Observed Values Versus Predicted Values of* $\log \hat{\lambda}_i$

Test number (index i)	Observed value of $\log \hat{\lambda}_i$	Predicted value of $\log \bar{\lambda}_i$	95 percent prediction limits	
			Lower	Upper
1	− 17.730	− 18.659	− 21.390	− 15.927
2	− 17.141	− 16.962	− 19.809	− 14.115
3	− 20.394	− 18.450	− 20.038	− 16.862
4	− 18.137	− 18.252	− 21.419	− 15.085
5	− 20.484	− 20.562	− 23.258	− 17.865
6	− 20.046	− 19.634	− 21.470	− 17.799
7	− 17.019	− 18.379	− 20.279	− 16.478
8	− 17.407	− 17.430	− 20.236	− 14.623

APPENDIX

RESULTS OF THE MONTE CARLO EXPERIMENTS

In this appendix we summarize the results of two Monte Carlo experiments that were performed in connection with this problem. Our first experiment is to test the approximate normality of $\log \hat{\lambda}_i$ and the second experiment is to estimate the actual coverage probabilities of the prediction intervals given by Equation (3.3). Since the two experiments complement each other, in the sense that one reaffirms the result of the other, the amount of data simulated in each experiment is not very large.

A.1. Test for Normality

We first choose the following values of r_i and λ_i:

$$r_i = 3, 5, 8, 11, 13, 15, 17, \text{ and } 19$$

$$\lambda_i = 0.001, 0.003, 0.005, 0.007, 0.010, 0.015, 0.030, 0.060, 0.100 \text{ and } 0.250.$$

We then generate 30 random numbers from a gamma distribution with a shape parameter r_i and a scale parameter $r_i\lambda_i$. We denote these random numbers by $\hat{\theta}_{ij}$, $j = 1, 2, \ldots, 30$; if we set $\hat{Y}_{ij} = \log \hat{\theta}_{ij}^{-1}$, then the distribution of $\hat{Y}_{ij}$ is of interest to us. We repeat this procedure for the 80 possible combinations of r_i and λ_i; thus, $l = 1, 2, \ldots, 80$.

All the 80 samples each comprised of 30 observations were tested for normality using the modified Kolmogorov-Smirnov test described in Lilliefors [4]. Since none of these were rejected, results of this simulation allow us to conclude the approximate normality of the $\hat{Y}_{ij}$, $j = 1, 2, \ldots, 30$, for each l, $l = 1, 2, \ldots, 80$, and thus the approximate normality of the $\log \bar{\lambda}_i$ of sections 2 and 3.

A.2. Coverage Probabilities of Prediction Intervals

The prediction intervals given in Equation (3.3) were calculated under the assumption that the $\log \hat{\lambda}_i$ are approximately normally distributed. Though this assumption was empirically verified in A.1. above, we (at the suggestion of a referee) simulated several accelerated life test experiments by using the Eyring Model and obtained prediction intervals for λ_u with the use of Equation (3.3) to see what

proportion of these prediction intervals covered the true parameter values. This Monte Carlo experiment and its results are summarized next.

1. We first choose A, B, C, D, V_u, and T_u, and compute λ_u by using Equation (2.1).

2. We next choose eight values of T and V each. Of these we select randomly a pair (T_i, V_i) $i=1, \ldots, 8$ and for each (T_i, V_i) pair, we compute λ_i by using Equation (2.1).

3. For each (T_i, V_i) pair we generate n_i random numbers from an exponential distribution with a parameter λ_i. We then compute $\hat{\lambda}_i$ using the methods of section 2 for some arbitrary value of r_i, $r_i \leq n_i$. We also compute z_i, where $z_i = \log(\hat{\lambda}_i / r_i T_i) - \psi(r_i)$, $i=1, 2, \ldots, 10$.

4. By using the results from Steps 2 and 3 above we obtain estimators of A, B, C, D, and λ_u with the use of Equations (3.1) and (3.2). We compute a 95-percent prediction interval for λ_u by using Equation (3.3) and note whether this interval contains the true value of λ_u obtained in 1 above.

5. We repeat Steps 2 through 4 for a total of 100 times, each time choosing different pairs of values of V_i and T_i, $i=2, \ldots, 8$.

6. We note the proportion of times the prediction interval contains λ_u.

7. We repeat Steps 1 through 6, nine more times, each time choosing a different set of values of A, B, C, D, V_u, and T_u.

Of the 10 times that the experiment was run, the proportion of times that the 95-percent prediction intervals cover the true value of λ_u is indicated below:

[Percent]

Run number:	1	2	3	4	5	6	7	8	9	10
Percentage of times the 95-percent prediction interval covers the true value of λ_u	98	96	97	95	94	96	98	97	95	96

The above results reaffirm our assumption regarding the approximate normality of $\log \lambda_i$.

REFERENCES

[1] Abramowitz, M. and I. Stegun, (eds.) *Handbook of Mathematical Functions* (Dover Publications, Inc., New York, 1965).

[2] Epstein, B. and M. Sobel, "Some Theorems Relevant To Life Testing From An Exponential Distribution," Ann. Math. Statist. *25*, 373–381 (1954).

[3] Graybill, F., *An Introduction To Linear Statistical Models* (McGraw-Hill, New York, 1961), vol. 1.

[4] Lilliefors, H., "On The Kolmogorov-Smirnov Test For Normality With Means And Variance Unknown," J. Amer. Statist. Assoc. *62*, 399–402 (1967).

[5] McGraw-Hill Encyclopedia Of Science And Technology (1966), vol. 7, p. 357.

[6] Pieruschka, E., "Relation Between Lifetime Distribution And The Stress Level Causing Failure," LMSD–800440, Lockheed Missiles and Space Division, Sunnyvale, California.

[7] Shooman, M. L., *Probabilistic Reliability: An Engineering Approach* (McGraw-Hill, New York, 1968).

[8] Singpurwalla, N. D., "Inference From Accelerated Life Tests When Observations Are Obtained From Censored Samples," Technometrics *13*, 161–170 (1970).

[9] Singpurwalla, N. D., "A Problem In Accelerated Life Testing," J. Amer. Statist. Assoc. *66*, 841–845 (1971).

[10] Singpurwalla, N. D., "Inference From Accelerated Life Tests Using Arrhenius Type Re-parameterizations," Technometrics *15*, 289–299 (1973).

[11] Thomas, R. (1964). "When Is A Life Test Truly Accelerated?" Electron. Design (Jan. 6, 1964).

AN OPERATOR THEORY OF PARAMETRIC PROGRAMMING FOR THE GENERALIZED TRANSPORTATION PROBLEM—III—WEIGHT OPERATORS*

V. Balachandran

Northwestern University
Evanston, Illinois

and

G. L. Thompson

Carnegie-Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

This paper investigates the effect on the optimum solution of a capacitated generalized transportation problem when any coefficient of any row constraint is continuously varied as a linear function of a single parameter. The entire analysis is divided into three parts. Results are derived relative to the cases when the coefficient under consideration is associated to a cell where the optimal solution in that cell attains its lower bound or its upper bound. The discussion relative to the case when the coefficient under consideration is associated to a cell in the optimal basis is given in two parts. The first part deals with the primal changes of the optimal solution while the second part is concerned with the dual changes. It is shown that the optimal cost varies in a nonlinear fashion when the coefficient changes linearly in certain cases. The discussion in this paper is limited to basis-preserving operators for which the changes in the data are such that the optimum bases are preserved. Relevant algorithms and illustrations are provided throughout the paper.

1. INTRODUCTION

In this paper we shall investigate the development of an "Operator Theory" of parametric programming for the generalized transportation problem for the case of weight operators, that was introduced in section 4 of [24]. The reader is assumed to be fully familiar with this reference [24] before attempting a study of this paper. In addition the reader is also assumed to be familiar with the paper [1], where the operator theory, as applied to cost, rim, and bounds for the generalized transportation problem is given.

We will consider operators which are basis preserving (Definition 15 in [24]) i.e., the original problem P and the transformed problem P^T share a common basis structure (basis B and sets LB and UB of cells at lower and upper bounds) in their (basic) optimal solutions. In section 2, we will discuss basis preserving weight-operators δE^+_{pq} and δE^-_{pq}, where (p, q) is not contained in the basis B.

*This report was prepared as part of the activities of the Management Science Research Group, Carnegie-Mellon University, Under Contract N00014-67-A-0314-0007 NR 047-048 with the U.S. Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

In section 3 we will consider the changes in the primal variables x_{ij} when δE^{+}_{pq} or δE^{-}_{pq} is applied with $(p, q)\epsilon B$. Section 4 deals with changes in the dual variables u_i and v_j due to weight operators $\delta E^{\pm}_{pq}$ when $(p, q)\epsilon B$ and finally summarizes the total changes in both primal and dual variables. Extension of cases when the basis preserving requirement is relaxed is considered in another paper [3]. In the present paper we shall use the same notation, bibliography, and examples as in [1, 24] and will continue numbering of theorems, remarks, equations, algorithms, figures, etc. from the paper [1].

2. BASIS PRESERVING WEIGHT OPERATORS ON NONBASIC CELLS

Let us now consider weight operators that transform the optimal solution when only the weight elements e_{ij} are changed in such a way that the optimal basis is preserved. To start with, we will consider a cell weight operator $\delta E^{\pm}_{pq}$ where e_{pq} is changed to $e_{pq}\pm\delta$ when $(p, q)\epsilon B$. Two cases arise; viz., when (i) $(p, q)\epsilon LB$ and (ii) $(p, q)\epsilon UB$. The following two theorems 13 and 14 give the required results.

ASSUMPTION A4: Hereafter, we will assume that $e_{ij} > 0$ for $i\epsilon I', j\epsilon J'$ to be meaningful in the context of a machine loading model. (However, the theoretical results derived later, can be easily modified if $e_{ij} \leqslant 0$.

THEOREM 13:

(a) For the operator δE^{+}_{pq} with $(p, q)\epsilon LB$,

$$D^{+} = D, \quad X^{+} = X, \quad Z^{+} = Z, \quad \text{and} \quad \mu^{+} = M, \tag{50}$$

(where M is a large positive number).

(b) For the operator δE^{-}_{pq} with $(p, q)\epsilon LB$,

$$D^{-} = D, \quad X^{-} = X, \quad Z^{-} = Z, \quad \text{and} \quad \mu^{-} = (c_{pq} - e_{pq}u_p - v_q)/|u_p|. \tag{51}$$

PROOF:

(a) Because $(p, q)\notin B$, $e^{+}_{ij} = e_{ij}$ for $(i, j)\epsilon B$. Thus for $(i, j)\epsilon B$ $e^{+}_{ij}u_i + v_j = e_{ij}u_i + v_j = c_{ij} = c^{+}_{ij}$. (Also $c^{\pm}_{ij} = c_{ij}$ because no change is made in costs.) Thus u^{+}_i and v^{+}_j are equal to u_i and v_j, respectively, for all $i\epsilon I$ and $j\epsilon J'$. Thus $D^{+} = D$. Since the operation is basis preserving $B^{+} = B$, $LB^{+} = LB$, $UB^{+} = UB$. Since

$$(p, q)\epsilon LB, \qquad Z^{+} = \sum_{i\epsilon I}\sum_{j\epsilon J'} c^{+}_{ij}x_{ij} = \sum_{i\epsilon I}\sum_{j\epsilon J} c_{ij}x_{ij} = Z.$$

Because the solution is both primal and dual feasible it is optimal.

The maximum extent that δE^{+}_{pq} can be applied is M where M is a large finite number. Since $(p, q)\epsilon LB$ by (18), $e_{pq}u_p + v_q \leqslant c_{pq}$. But at the optimal solution all $u_i \leqslant 0$ for $i\epsilon I$. Thus $(e_{pq} + \delta)u_p + v_q = (e_{pq}u_p + v_q) + \delta u_{pq} \leqslant c_{pq} + \delta u_p \leqslant c_{pq}$ for $\delta > 0$. The last inequality holds because $u_p \leqslant 0$. see Equation (16). Thus $\mu^{+} = M$.

(b) Proof of this part is similar.

REMARK 17: Let us apply an area operator δE^{A} such that $e^{A}_{ij} = e_{ij} + \delta\epsilon_{ij}$ for $i\epsilon I', j\epsilon J'$ with the requirement that $\epsilon_{ij} = 0$ for $(i, j)\notin LB$ and ϵ_{ij} is arbitrary for $(i, j)\epsilon LB$. It is easy to see from Theorems 13(a) and 13(b) that $D^{A} = D$, $X^{A} = X$, $Z^{A} = Z$ and μ^{A} is given to be

$$\mu_A = \underset{\{i|e_{ij}<0\}}{\text{Min}} \left\{ \frac{(c_{ij} - e_{ij}u_i - v_j)}{u_i e_{ij}} \right\}. \tag{52}$$

Again μ^- given in Theorem 13(b) is defined to preserve the dual feasibility only. It is quite likely that the new $e_{pq}^- = e_{pq} - \mu^-$ may become negative. Since we want to preserve $e_{pq} > 0$ then $\delta < \mu^-$ should be chosen so that $e_{pq}^- > 0$.

Now consider the case of δE_{pq}^+ with $(p, q) \epsilon UB$. Here, the primal feasibility for row p will be violated since $a_p^+ = a_p + \delta U_{pq} \cdot (x_{pq} = U_{pq})$. To restore primal feasibility for just the row p we need a new type of rim operator and hence the following specific rim operator is given.

REMARK 18: Consider $\delta R_{p,n+1}^+$ $(\delta R_{p,n+1}^-)$, when the row total $a_p^+ = a_p + \delta$; $(a_p^- = a_p - \delta)$ and $a_i^+ = a_i (= a_i^-) i \neq p$. Also $b_j^+ = b_j (= b_j^-)$. (Note $e_{p,n+1} \equiv 1$ by definition and $(n+1)$th column being slack is unconstrained.) Thus the results of Theorems 3–5 are applicable to this special type of operator. We will give two theorems for $\delta E_{pq}^{\pm}$ with $(p, q) \epsilon UB$ when $(p, n+1) \epsilon B$ (Theorem 14) and when $(p, n+1) \epsilon B$ (Theorem 15).

THEOREM 14:

(a) For the operator δE_{pq}^+ with $(p, q) \epsilon UB$ and $(p, n+1) \epsilon B$, the optimal primal solution X^+ to the problem P^T is given by

$$\begin{cases} x_{ij}^+ = x_{ij} & \text{for } (i, j) \epsilon [(IxJ) - \{(p, n+1)\}] \\ x_{p,n+1}^+ = x_{p,n+1} - \delta U_{pq}. \end{cases} \tag{53}$$

Further, $D^+ = D$; $Z^+ = Z$. The maximum extent to which this preserving operator can be applied is given by

$$\mu^+ = x_{p,n+1}/U_{pq}. \tag{54}$$

(b) For the operator δE_{pq}^- with $(p, q) \epsilon UB$ and $(p, n+1) \epsilon B$, the optimal primal solution X^- to the problem P^T is given by

$$\begin{cases} x_{ij}^- = x_{ij} & \text{for } (i, j) \epsilon [(IxJ) - \{(p, n+1)\}] \\ x_{p,n+1}^- = x_{p,n+1} + \delta U_{pq}. \end{cases} \tag{55}$$

Further $D^+ = D$; $Z^+ = Z$. The maximum extent to which this basis preserving operator can be applied is given by

$$\mu^- = e_{pq}. \tag{56}$$

PROOF:

(a) Since $(p, q) \epsilon UB$, following the arguments of Theorem 13(a), $e_{ij}^+ = e_{ij}$ for $(i, j) \epsilon B$. Hence, it is clear that $D^+ = D$. Since $x_{pq} = U_{pq}$, due to operator δE_{pq}^+, $a_p^+ = a_p + U_{pq}\delta$, so that pth row rim con-

straint is violated. Thus to problem P^T let us now apply the cell rim operator (of remark 18) $\delta' R^-_{p,n+1}$, where $\delta' = \delta U_{pq}$. Since we are given $(p, n+1) \epsilon B$, Theorem 4(b) [1] is applicable which yields $x^-_{p,n+1} = x_{p,n+1} - \delta'$ and $x^-_{ij} = x_{ij}$ for $(i, j) \epsilon [(IxJ)] - \{(p, n+1)\}$. The maximum extent μ^- for δ' is $x_{p,n+1}$. Since $\delta' = \delta U_{pq}$, incorporating these results due to $\delta' R^-_{p,n+1}$ to the original problem P via P^T, it is easy to see that X^+ to P^T is the same as given by (53) and $\mu^+ = x_{p,n+1}/U_{pq}$. From Theorem 3(b) $Z^- = Z - \delta' d_{p,n+1}$. But $d_{p,n+1} = e_{p,n+1} u_p = 0$, since $u_p = 0$ at the optimum. Since Z^- of $\delta' R^-_{p,n+1}$ is equal to Z^+ of δE^+_{pq}, $Z^+ = Z$.

(b) Proof for part (b) is similar to that for (a) except the following. While determining μ^- for δE^-_{pq}, we look at the maximum extent for the operator $\delta' R^+_{p,n+1}$. This from Theorem 4(a) [1] is $(U_{p,n+1} - x_{p,n+1}) = M$ a large positive number. Thus μ^- for $\delta E^-_{pq} = M/U_{pq} = M'$ another large positive number. However, since we assume all e_{ij} to be > 0, we give the maximum extent μ^- as e_{pq} in (56).

THEOREM 15:

(a) For the operator δE^+_{pq} with $(p, q) \epsilon UB$ and $(p, n+1) \epsilon B$, the optimal primal solution X^+ to the problem P^T is given by

$$x^+_{ij} = \begin{cases} x_{ij} + m_{ij} \delta U_{pq} & \text{for } (i, j) \epsilon \Gamma - \{(p, n+1)\} \\ x_{ij} & \text{elsewhere,} \end{cases} \tag{57}$$

where Γ is the row segment (since the column segment of Γ is $\emptyset$) defined in Remark 7 [24], and m_{ij}'s are given by algorithm A4(b). Further $D^+ = D$ and $Z^+ = Z - \delta/U_{pq} u_p$. The maximum extent to which this basis preserving operator can be applied is given by

$$\mu^+ = \text{Min} \begin{cases} x_{ij}/(-m_{ij} U_{pq}) & \text{for } (i, j) \epsilon \Gamma_2 \\ (U_{ij} - x_{ij})/(m_{ij} U_{pq}) & \text{for } (i, j) \epsilon \Gamma_1 - \{(p, q)\}, \end{cases} \tag{58}$$

where Γ_1 and Γ_2 are given by algorithm A4(b).

(b) For the operator δE^-_{pq} with $(p, q) \epsilon UB$ and $(p, n+1) \notin B$, the optimal solution X^- to the problem P^T is given by

$$x^-_{ij} = \begin{cases} x_{ij} + m_{ij} \delta U_{pq} & \text{for } (i, j) \epsilon \Gamma - \{(p, n+1)\} \\ x_{ij} & \text{elsewhere,} \end{cases} \tag{59}$$

where m_{ij}'s are given by A4(a). (Note the m_{ij}'s given in (59) are the negatives of m_{ij}'s given in (57)). Further $D^- = D$ and $Z^- = Z + \delta U_{pq} u_p$. The maximum extent to which this basis-preserving operator can be applied is given by

$$\mu^- = \text{Min} \begin{cases} (U_{ij} - x_{ij})/(m_{ij} U_{pq}) & \text{for } (i, j) \epsilon \Gamma_1 - \{(p, q)\} \\ x_{ij}/(-m_{ij} U_{pq}) & \text{for } (i, j) \epsilon \Gamma_2. \end{cases} \tag{60}$$

(Note Γ_1 and Γ_2 are different from those given in (58) and given by algorithm A4(a) here.)

PROOF:

(a) The proof of this part is similar to that given for Theorem 14(a). However, while we apply the cell rim operator $\delta' R^{-}_{\bar{p},n+1}$ where $\delta' = \delta U_{pq}$, we note that $(p, n+1) \notin B$. Hence, we invoke Theorem 5(b) [1] for the case $(p, n+1) \notin B$. Since the two tree Γ of definition 10 [24], due to the introduction of $(p, n+1)$ does not possess a column segment (col. $(n+1)$ is never connected), the two tree Γ becomes just the row segment. The rest of the proof is similar to the proof of Theorem 5(b) [1].

(b) Proof for part (b) is similar.

EXAMPLE (Continued from [1]): Consider Figure 17, where δE^{+}_{12} is applied where $(1, 2) \epsilon LB$. It is easy to see that $\mu^{+} = M$ while preserving the basis and optimality. In fact if we apply an area operator with $\epsilon_{ij} \geq 0$ for $(i, j) \epsilon \{(1, 2), (2, 3), (2, 4), (3, 3), (3, 4)\}$ and $\epsilon_{ij} = 0$ elsewhere, we can see in this case $\mu^{A} = M$. However, if we apply $\delta E^{-}_{\bar{2}\bar{3}}$ then $\mu^{-} = (c_{23} - e_{23}u_2 - v_3)/|u_2| = (8 - 5(-1/5) - 3)/(1/5) = 30$. However, from Remark 16, we must keep $\delta < 5$ so that $e_{pq} > 0$.

u_i \ v_j →	31/5	28/5	3	4	0	a_i
0	3 15 2 15	2+δ 6 N	2 35 (3) 40	3 60 (4) 80	1 5 (0) N	300
−1/5	1 135 (6) 150	3 30 (5) 75	5 8 25	10 15 N	1 0 N	225
−13/10	4 20 (1) 60	2 30 (3) N	2 10 30	1 6 50	1 0 N	140
0	1 100 N	1 100 N	1 100 N	1 100 N	1 1000 (0) N	1000
b_j	170	60	35	60	—	

FIGURE 17. δE^{+}_{12} $(0 \leq \delta \leq M)$.

Consider Figure 18, where δE^{+}_{11} is applied. Here $(1, 1) \epsilon UB$ and $(1, 5) \epsilon B$. Thus Theorem 14(a) is applicable. Since $e^{+}_{11} = e_{11} + \delta$, the first row total will be $a^{+}_{1} = 300 + 15\delta$. Hence, to restore primal feasibility for row 1, the cell rim operator $\delta' R^{-}_{\bar{p},n+1} = 15\delta R^{-}_{1,5}$ is applied. Due to (54) of Theorem 14(a) $\mu^{+} = x_{15}/U_{11} = 1/3$. Thus for $0 \leq \delta \leq \mu^{+}$ the basis preserving property is preserved. The cost change will be 0, $(Z^{+} = Z)$ and $D^{+} = D$. The optimal primal solution X^{+} to the problem P^{T} due to (53) is given below:

$$x^{+}_{15} = x_{15} - 15\delta = 0$$

$$x^{+}_{ij} = x_{ij} \qquad \text{for all other } (i, j)$$

u_i \ v_j →	31/5	28/5	3	4	0	a_i
0	3+δ 15 3 15	2 6 N	2 35 (3) 40	3 60 (4) 80	1 0 (0) N	300
−1/5	1 135 (6) 150	3 30 (5) 75	5 9 25	10 15 N	1 0 N	225
−13/10	4 20 (1) 60	2 30 (3) N	2 10 30	1 6 50	1 0 N	140
0	1 100 N	1 100 N	1 100 N	1 100 N	1 1000 (0) N	1000
b_j	170	60	35	60	—	

FIGURE 18. δE_{11}^{+} $(0 \leq \delta \leq 1/3)$.

3. BASIS PRESERVING WEIGHT OPERATORS ON BASIC CELLS: PRIMAL CHANGES

Let us now consider weight operator $\delta E_{pq}^{\pm}$ when (p, q) is in the basis. In this section, we will restrict our attention only to changes in the primal variables x_{ij}, where $(i, j) \epsilon B$. Thus, we will ensure that the transformed primal problem P^T is primal feasible. For all the different cases to be considered hereafter, we will provide expressions which will give the changes in the primal variables due to the application of $\delta E_{pq}^{\pm}$. We will give algorithms for computing the changes in the primal variables. Moreover, the maximum extent to which an operator can be applied while maintaining both primal feasibility and the basis preserving property will be given. This extent will be called the *maximum primal extent* and will be denoted by the symbol $\mu_P^{\pm}$ (with a subscript P). Thus there is no guarantee that the transformed problem P^T is optimal at this stage because we have not considered dual feasibility. However in section 4 we will develop analogous cases for the changes in the dual variables due to application of $\delta E_{pq}^{\pm}$ and will provide algorithms. In addition, we will find the *maximum dual extent*, $\mu_D^{\pm}$ which will give the maximum extent the operator can be applied satisfying dual feasibility (17)–(19) and the basis preserving property. Finally combining these two sets of results and defining the maximum extent $\mu^{\pm} = \min \{\mu_P^{\pm}, \mu_D^{\pm}\}$, we can see that for $\delta \leq \mu^{\pm}$ the transformed problem P^T is both primal and dual feasible and hence optimal.

REMARK 19: As in the case of the cost operator [1], we will classify four mutually exclusive and collectively exhaustive cases with $G(T_k) = \{I_k \cup J_k, B_k\}$

$$
\text{(61)} \quad \begin{cases} \text{(1)} & (p, q) \epsilon C_k & \text{and} & p_k(p) = q; & \\ \text{(2)} & (p, q) \epsilon C_k & \text{and} & p_k(q) = p; & \\ \text{(3)} & (p, q) \epsilon C_k^* & \text{and} & p_k(p) = q; & \text{and} \\ \text{(4)} & (p, q) \epsilon C_k^* & \text{and} & p_k(q) = p. & \end{cases}
$$

We will develop algorithms A12(a), (b), (c) which will provide the primal changes λm_{ij}^{+} for x_{ij} with $(i, j)\epsilon B_k$ when $\delta E_{pq}^{\pm}$ is applied. Though these algorithms are for δE_{pq}^{+}, those for δE_{pq}^{-} are identical, except in Step (0) we put $-\delta$ and δ and interchange the sets Γ_1 and Γ_2 in A12(b) and (c). (The case of $\delta E_{i,n+1}^{\pm}$ does not arise because $e_{i,n+1} \equiv 1$ and cannot be changed.)

ALGORITHM A12(a). For finding the primal changes $m_{ij}\lambda$, for all $(i, j)\epsilon C_k$ when the operator δE_{ef}^{+} is applied given $(e, f)\epsilon C_k$ (cases (1) and (2) in (61)).

(0) Denote C_k by C. Let $(u, v) = (e, f)$. (Note $(p, q) = (e, f)$). If $p(u) = v$, set $\lambda = (\alpha\delta x_{uv})/[e_{uv}(\beta-\alpha)-\alpha\delta]$. Let $m_{uv}=-1$, $(u, v)\epsilon\Gamma_2$ and go to (1). If $p(v)=u$, let $\lambda=-(\beta\delta x_{uv})/[e_{uv}(\alpha-\beta)-\beta\delta]$. Let $m_{uv}=-1$ and $(u, v)\epsilon\Gamma_1$ if $e_{uv}(\alpha-\beta) \geq \beta\delta$. Else $(u, v)\epsilon\Gamma_2$ and go to (2).

(1) Let $i = p(v)$. Set $m_{iv} = -m_{uv}$. If $\lambda m_{iv} \geq 0$ let $(i, v)\epsilon\Gamma_1$, else $(i, v)\epsilon\Gamma_2$. If $i = e$ go to (3). Else let $u = i$.

(2) Let $j=p(u)$. Set $m_{uj}=-e_{uv}m_{uv}/e_{uj}$. If $\lambda m_{uj} \geq 0$, let $(u, j)\epsilon\Gamma_1$ else $(u, j)\epsilon\Gamma_2$. If $j=f$ go to (3). Else let $v = j$ and go to (1).

(3) Go to Algorithm A12(c) for finding μ_{λ}^{+} and x_{ij}.

REMARK 20: Consider the case $\delta E_{ef}^{\pm}$, where $(e, f)\epsilon C^*$. Let θ_k represent the path connecting row e, if $p(e) = f$, (or column f if $p(f) = e$) to a node $w\epsilon S(C_k)$ (i.e., $d(w) = 0$) such that $l(e, w)$ $(l(f, w))$ is minimum. Then every cell $(i, j)\epsilon(\theta_k \cup C_k)$ will have a change in x_{ij} and all other x_{ij} such that $(i, j)\epsilon B-\{\theta_k \cup \bar{C}_k\}$ will not change. Further, if $(e, f)\epsilon C_k$ then $\theta_k = \phi$. Algorithm A12(c) will give x_{ij}^{+} for all $(i, j)\epsilon\theta_k \cup C_k$ for the case of δE_{ef}^{+}. Algorithm A12(b) will provide the factors m_{ij} for $(i, j)\epsilon\theta_k \cup C_k$ which are needed for A12(c). Since C_k can be either a cycle or a loop, Algorithm A12(b) considers both cases.

ALGORITHM A12(b). For finding primal changes $m_{ij}\lambda$ for all $(i, j)\epsilon\theta_k \cup C_k$ when the operator δE_{ef}^{+} is applied given $(e, f)\epsilon C_k^*$ (cases (3) and (4) in (61)).

(0) Denote C_k, θ_k and C_k^* by C, θ and C^*, respectively. Let $(u, v) = (e, f)$. If $p(u) = v$ set $\lambda = \delta x_{uv}/(e_{uv}+\delta)$ and $m_{uv}=-1$; let $(u, v)\epsilon\Gamma_2$ and go to (1). If $p(v) = u$ set $\lambda = \delta x_{uv}$ and $m_{uv} = 0$. Let $j = p(u)$ and set $m_{uj} = -1/e_{uj}$. Let $(u, j)\epsilon\Gamma_2$, and $v = j$. Set $i^* = j^* = \phi$.

(1) Let $i = p(v)$. Set $m_{ij} = -m_{uv}$. Let $(i, v)\epsilon\Gamma_1$. Let $u = i$. If $d(u) = 0$ go to (3). Else go to (2).

(2) Let $j = p(u)$. Set $m_{uj} = -e_{uv}m_{uv}/e_{uj}$. Let $(i, j)\epsilon\Gamma_2$ and $v = j$. If $d(v) = 0$ go to (4). Else go to (1).

(3) Let $j = p(u)$. If $j = n + 1$ set $m_{u,n+1} = -e_{uv}m_{uv}$. Let $(u, n + 1)\epsilon\Gamma_2$ and go to (7). Else set $m_{uj} = -\rho e_{uv}m_{uv}/e_{uj}$. Let $(i, j)\epsilon\Gamma_2$ and $v = j$ and $j^* = j$ and go to (5).

(4) Let $i = p(v)$. Set $m_{iv} = -\rho m_{uv}$ and $(i, v)\epsilon\Gamma_1$. Let $u = i$, $i^* = i$ and go to (6).

(5) Let $i = p(v)$. Set $m_{iv} = -m_{uv}$ and $(i, v)\epsilon\Gamma_1$. If $i = i^*$ go to (7). Else let $u = i$.

(6) Let $j=p(u)$, set $m_{uj}=-e_{uv}m_{uv}/e_{uj}$ and $(u, i)\epsilon\Gamma_2$. If $j=j^*$ go to (7). Else let $v=j$ and go to (5).

(7) Go to Algorithm A12(c) for finding μ_{λ}^{+} and x_{ij}.

ALGORITHM A12(c). For finding new primal variables x_{ij}^{+} for $(i, j)\epsilon\theta_k \cup C_k$ given m_{ij} and λ (A12(a), (b)) when δE_{ef}^{+} is applied where $(e, f)\epsilon B_k$.

(1) Compute the maximum value for λ from

$$\mu_\lambda^+ = \text{Min} \begin{cases} (U_{ij} - x_{ij})/m_{ij} & \text{for } (i,j)\epsilon\Gamma_1 \\ x_{ij}/(-m_{ij}) & \text{for } (i,j)\epsilon\Gamma_2. \end{cases} \tag{62}$$

The cell on which the maximum is taken is the leaving cell (e_1, f_1). If $(e_1, f_1)\epsilon\Gamma_1$ then this cell is in the basis at its maximum level and if $(e_1, f_1)\epsilon\Gamma_2$ it is in the basis at zero level.

(2) Compute the maximum extent μ_P^+ for δ as follows: $(0 \leqslant \delta \leqslant \mu_P^+)$

$$\begin{cases} \mu_P^+ = \text{Max} \begin{cases} 0 \\ \dfrac{[e_{ef}(\beta - \alpha)\mu_\lambda^+]}{[\alpha(\mu_\lambda^+ - x_{ef})]} \end{cases} & \text{for case (1) of (61)} \\ \mu_P^+ = \text{Max} \begin{cases} 0 \\ \dfrac{[e_{ef}(\alpha - \beta)\mu_\lambda^+]}{[\beta(\mu_\lambda^+ - x_{ef})]} \end{cases} & \text{for case (2) of (61)} \\ \mu_P^+ = \text{Max} \begin{cases} 0 \\ e_{ef}\mu_\lambda^+/(x_{ef} - \mu_\lambda^+) \end{cases} & \text{for case (3) of (61)} \\ \mu_P^+ = \mu_\lambda^+/x_{ef} & \text{for case (4) of (61)} \end{cases} \tag{63}$$

(Note: If $\mu_\lambda^+ = x_{ef}$ in the first three cases, set $\mu_P^+ = M$ where M is a large positive number.)

(3) Set

$$x_{ij}^+ = \begin{cases} x_{ij} + \lambda m_{ij} & \text{for } (i,j)\,\epsilon\,\theta_k \cup C_k \\ x_{ij} & \text{otherwise} \end{cases} \tag{64}$$

STOP.

REMARK 21: In step (2) of Algorithm A12(c) the maximum primal extent $\mu_P^\mp$ in terms of $\mu_\lambda^\mp$ is computed. Since the changes in x_{ij} are functions of $\mu_\lambda^\mp$ we compute this first. Moreover λ also measures the change in Z (68) so that λ itself has managerial interpretations. It is easy to see that the relationship of λ is nonlinear in δ for the first three cases while it is linear in δ for the fourth case. Hence in the first three cases, if we express $\lambda = \delta/(A - B\delta)$ where A and B are constants (in terms of α, β, e_{ef} and x_{ef}), then $\delta = A/B$ makes λ infinite (i.e., $\delta = A/B$ is a pole for λ). Thus for $\delta \doteq A/B$, x_{ef} is to be changed to ∞. This is obviously impossible in practice. In [3] we will discuss, for specific examples, the managerial interpretations of δ, λ, and the weight operators.

PROPOSITION 3: For the operator $\delta E_{pq}^\mp$ with $(p, q)\epsilon B_k$

$$x_{ij}^{\pm} = \begin{cases} x_{ij} + m_{ij}\lambda & \text{for } (i,j) \in C_k \cup \theta_k \\ x_{ij} & \text{elsewhere} \end{cases} \tag{65}$$

where C_k is the cycle of B_k and θ_k is the path given in Remark 20, m_{ij} and λ are given by Algorithm A12. The maximum extent for λ to be basis preserving is given by

$$\mu_{\lambda}^{\pm} = \text{Min} \begin{cases} (U_{ij} - x_{ij})/m_{ij} & \text{for } (i,j) \in \Gamma_1 \\ x_{ij}/(-m_{ij}) & \text{for } (i,j) \in \Gamma_2, \end{cases} \tag{66}$$

where Γ_1 and Γ_2 are given by Algorithm A12, such that $\Gamma_1 \cup \Gamma_2 = C_k \cup \theta_k$. The maximum primal extent of δ is given by

$$\mu_P^{\pm} = f(\mu_{\lambda}^{\pm}), \tag{67}$$

where the functional expression is given in Equation (63). The change in total cost will be

$$Z^{\pm} - Z = \lambda \left[\sum_{(i,j) \in \Gamma_1 \cup \Gamma_2,} c_{ij} m_{ij}^{\pm} \right] \tag{68}$$

where λ is given in step (0) of algorithms A12(a) or A12(b) and $m_{ij}^{\pm}$ are those obtained in algorithm A12 with $\pm\delta$ respectively in Step (0)

PROOF: We will give the proof for δE_{pq}^{+} for the four cases given in (61). The proof for δE_{pq}^{-} is similar. Let $(e, f) = (p, q)$ so that the algorithms A12(a) and (b) can be used and consider case (1) where $p(e) = f$ and $(e, f) \in C_k$. In this case $C_k \cup \theta_k = C_k$. Since $p(e) = f$ there is one more cell (i, f) in column f. Moreover $(e, f) \in S_\alpha$. Let $x_{ef}^{+} = x_{ef} - \lambda$. Then to satisfy the column feasibility we have to make $x_{if}^{+} = x_{ij} + \lambda$, where $i = p(f)$. Thus traversing along the graph of the cycle C_k as dictated by the predecessor function, balancing alternatively column and then row requirements (thus preserving primal feasibility for every node in the span of C_k) we reach an edge (e, j), where $p(j) = e$. The basic variable x_{ej} corresponding to this edge (e, j) will be changed to

$$x_{ej}^{+} = x_{ej} + \lambda [(\beta/e_{ej})/\alpha/e_{ef})].$$

Since the primal constraint for row e is to be satisfied, we have

$$(e_{ef} + \delta)(x_{ef} - \lambda) + e_{ej}[x_{ej} + (\lambda \beta e_{ef}/\alpha e_{ej})] + A = a_e,$$

where $A = \sum_{v} e_{ev} x_{wv}$, $v \in \{J - f - j\}$. But from problem P we have

$$e_{ef} x_{ef} + e_{ej} x_{ej} + A = a_e.$$

Thus solving for λ from these two expressions given above we get the expression for λ as given in Algorithm A12(a). The m_{ij}'s given in A12(a) are obtained by the procedure described above. Since,

we assume that each $e_{ij} \geq 0$, $\alpha \geq 0$ and $\beta \geq 0$, and $(\alpha - \beta) \geq 0$ by definition 2, so that $\lambda \geq 0$. Since we put $x_{ef}^{+} = x_{ef} - \lambda$ we give $m_{ef} = -1$ so that when $\lambda \geq 0$ $(e, f) \in \Gamma_2$, the set of all negative coefficients. Thus the maximum extent this λ can take without changing the basis structure is given by Equation (62). The x_{ij}^{+} given in (64) preserve the primal feasibility conditions of (6) and (7) for the nodes contained in $S(C_k)$. Since $x_{ij}^{+} = x_{ij}$ for $(i, j) \notin C_k$, the primal feasibility conditions of (6) and (7) are satisfied for the problem P^T. Substituting the value of μ_{λ}^{+} for λ we can solve for δ which is given by μ_{P}^{+} as shown in (63). Since the changes in x_{ij} are only for cells $(i, j) \in C_k$, $Z^{+} - Z$ depends on c_{ij} and the changes m_{ij} for $(i, j) \in C_k$. Since Algorithm A12(a) classifies cells $(i, j) \in C_k$ into Γ_1, where $m_{ij} \geq 0$ and to Γ_2 where $m_{ij} < 0$, it is easy to verify that $Z^{+} - Z$ as given by (68) is true.

The proof for case (2) of (61) where $p(f) = e$ and $(e, f) \in C_k$ is similar to that of case (1).

Let us now consider the proof for case (3) where $(e, f) \in C_k^{*}$ and $p(e) = f$. Since $p(e) = f$, (e, f) is the only cell $\in \theta_k$ in row e. To preserve the rim condition of row e, let λ be the change in x_{ef} such that $x_{ef}^{+} = x_{ef} - \lambda$. If λ is chosen to preserve the primal feasibility for row e, then $(e_{ef} + \delta)$ times $(x_{ef} - \lambda) = e_{ef} x_{ef}$, which gives $\lambda = \delta x_{ef} / (e_{ef} + \delta)$. Note that λ is positive and hence if $m_{ef} = -1$, then $x_{ef}^{+} \leq x_{ef}$ so that $(e, f) \in \Gamma_2$ (the set of cells (i, j) where $m_{ij} \leq 0$). Algorithm A12(b) provides the coefficients m_{ij} of λ for $(i, j) \in \theta_k \cup C_k$ such that $x_{ij}^{+} = x_{ij} + m_{ij}\lambda$ as given in (64). The rest of the proof is similar to the proof for case (1). It is to be noted that in these three cases ((1), (2), and (3) of (61)) λ is a nonlinear function of δ.

In case (4) we have $(e, f) \in \theta_k$ in row e, (where $j = p(e)$), which can balance the change in the rim condition for row e. Thus m_{ef} is set equal to zero in A12(b). Since $e_{ef}^{+} = e_{ef} + \delta$, a_e is increased to $a_e + \delta x_{ef}$. Thus this addition $\delta x_{ef} = \lambda$ can be adjusted by putting $x_{ej}^{+} = x_{ej} + m_{ej}\lambda$ with $m_{ej} = -1/e_{ei}$, where $j = p(e)$. Since $\lambda > 0$, $m_{ej} < 0$, so that $(e, j) \in \Gamma_2$ the set of cells (i, j) where $m_{ij} < 0$. The rest of the proof is similar to that for case (3). It is to be noted that only in this case λ is linear in δ.

The proof for δE_{pq}^{-} is similar for all four cases of (57).

EXAMPLE (continued): Consider Figure 19, where δE_{22}^{+} is applied with $(2, 2) \in B$. It is seen that

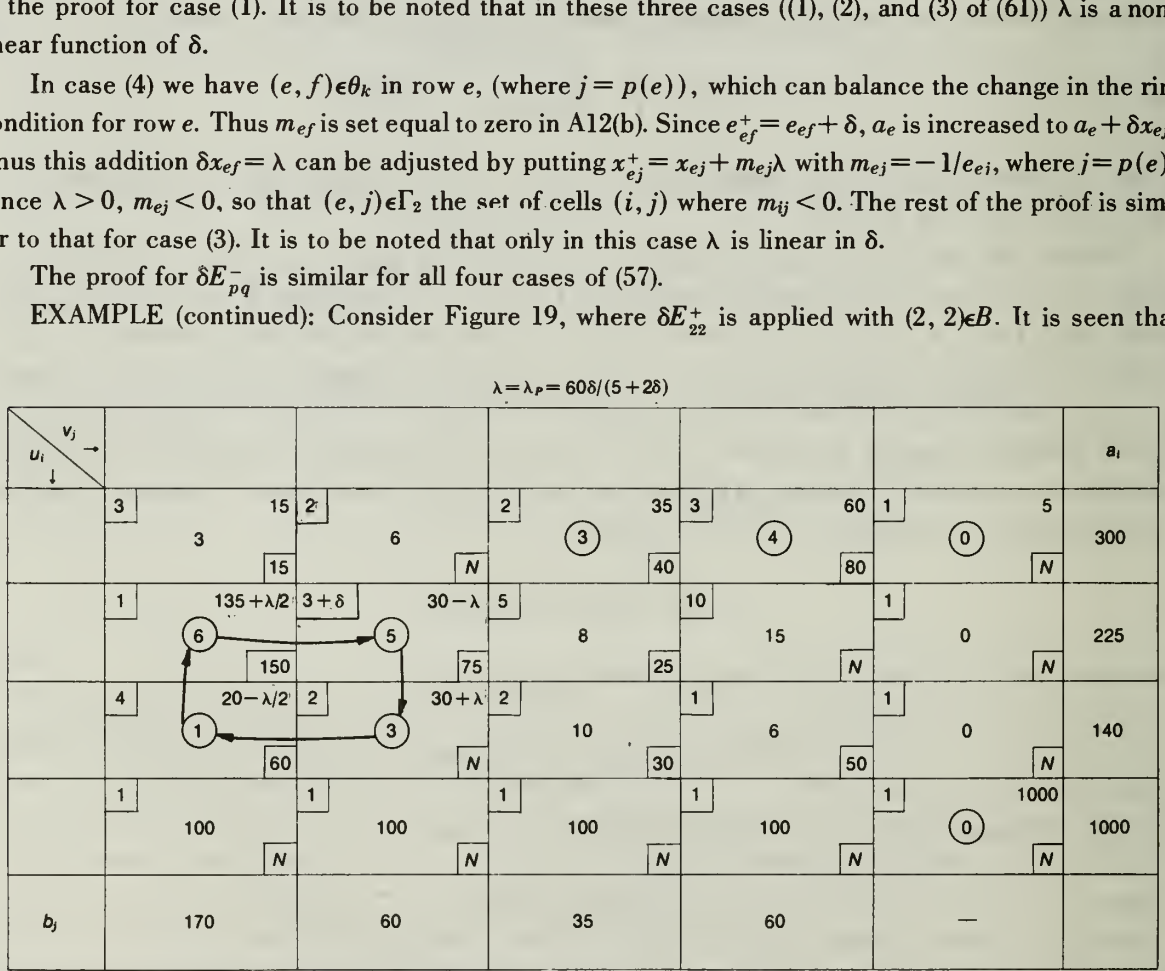


FIGURE 19. Primal changes due to δE_{22}^{+}.

$(2, 2)\epsilon C_2$ and $p(\text{row } 2) = \text{col. } 2$. (This predecessor relation is indicated by arrows.) Thus case (1) of (61) is applicable. Here $\alpha = 12$, $\beta = 2$. Let us apply Algorithm A12(a) and A12(c) to find the transformed primal variables and the maximum primal extent μ_p^+. The following gives the steps of A12(a) and A12(c) and their results.

Steps of A12(a)	*Result*
(0)	$(u, v) = (2, 2)$. Since $p(R_2) = C_2$ set $\lambda = -360\delta/[3(-10) - 12\delta] = 60\delta/(5 + 2\delta)$. Let $m_{22} = -1$ and $(2, 2)\epsilon\Gamma_2$. Go to (1).
(1)	$R_3 = p(C_2)$. Set $m_{32} = -m_{22} = 1$; $(3, 2)\epsilon\Gamma_1$, $u = R_3$.
(2)	$C_1 = p(R_3)$. $m_{31} = -2/4 = -1/2$; $(3, 1)\epsilon\Gamma_2$. $v = C_{11}$. Go to (1).
(1)	$R_2 = p(C_1)$. $m_{21} = -m_{31} = 1/2$; $(2, 1)\epsilon\Gamma_1$. Since $R_2 = e$, go to (3). (Go to A12(c)).

Steps of A12(c)	*Result*
(1)	$\mu_\lambda^+ = \text{Min} \begin{cases} (N-30)/1, & (150-135)/(1/2) \\ 30/1, & (20)/(1/2) \end{cases}$
	$= 30.$
	Leaving cell (e_1, f_1) is either $(2, 1)$ or $(2, 2)$. Assume that $(e_1, f_1) = (2, 1)$.
(2)	Since $\mu_\lambda^+ = x_{ef} = 30$, set $\mu_P^+ = M$ a large positive number.
(3)	$x_{22}^+ = x_{22} - \lambda$
	$x_{32}^+ = x_{32} + \lambda$
	$x_{31}^+ = x_{31} - \lambda/2$
	$x_{21}^+ = x_{21} + \lambda/2$,
	where $\lambda = 60\delta/(5 + 2\delta)$.

EXAMPLE (continued): Consider Figure 20, where let $\lambda = 20 < \mu_\lambda^+ = 30$. Since $20 = 60\delta/(5 + 2\delta)$, $\delta = 5$. Utilizing the results given by A12(a) and A12(c) given above, we have, X^+ for P^T will be

$$x_{22}^+ = x_{22} - \lambda = 10$$

$$x_{32}^+ = x_{32} + \lambda = 50$$

$$x_{31}^+ = x_{31} - \lambda/2 = 10$$

$$x_{21}^+ = x_{21} + \lambda/2 = 150$$

$$x_{ij}^+ = x_{ij} \qquad \text{for } (i, j)\,\epsilon B - C_2.$$

$$e_{22}^+ = e_{22} + \delta = 3 + 5 = 8.$$

These are given in Figure 20.

$\lambda = \lambda_p = 60\delta/(5+2\delta) = 20$

$u_i \downarrow$ \ $v_j \rightarrow$						a_i
	3 15 **3** 15	2 6 N	2 35 **3** 40	3 60 **4** 80	1 5 **0** N	300
	1 145 **6** 150	8 10 **5** 75	5 8 25	10 15 N	1 0 N	225
	4 10 **1** 60	2 50 **3** N	2 10 30	1 6 50	1 0 N	140
	1 100 N	1 100 N	1 100 N	1 100 N	1 1000 **0** N	1000
b_j	170	60	35	60	—	

FIGURE 20. Primal changes due to $5E^{+}_{22}$.

4. BASIS PRESERVING WEIGHT OPERATORS ON BASIC CELLS: DUAL CHANGES

Let us now consider the dual variables when the operator δE^{+}_{pq} is applied. Here also we will consider the same four cases as given in (61). However, the discussion will be very similar to that of the cost (cell) operator δC^{+}_{pq} [1] and hence we will define $\Omega_k = B_k - \{(p, q)\}$ and partition I and J' as I_p, I_q and J'_p, J'_q, respectively, as in (38) [1]. The dual changes here are similar to those of the cost operator and are given in Algorithms A13(a) and (b). (Similar to A9(a) and (b)). The arguments for these algorithms are the same as in the case of cost operators. Once again for algorithm for δE^{-}_{pq} statement similar to δC^{-}_{pq} [1] holds and thus the dual variables are given in (71).

ALGORITHM A13(a). For finding the dual variables u_i, v_j when the operator δE^{+}_{ef} is applied given that $(e, f) \epsilon C_k$ (cases (1) and (2) of (61)).

(0) Denote C_k by C and C^*_k by C^*. Let $(i, j) = (e, f)$. (Note that $(p, q) = (e, f)$). If $p(i) = j$, set $\lambda = -(\delta\beta u_i e_{ij})/[\alpha\delta + e_{ij}(\alpha - \beta)]$ (note that $\lambda \geq 0$, since $u_i < 0$) and $n_j = -1$. Let $v_j = v_j + n_j\lambda$ and go to (1). If $p(j) = i$ set $\lambda = -\delta\beta u_i/[\beta\delta + e_{ij}(\alpha - \beta)]$ (again $\lambda \geq 0$ due to $u_i \leq 0$) and $m_i = -1$. Let $u_i = u_i + m_i\lambda$ and go to (2).

Steps (1), (2), and (3) are same as those in A9(a) [1], where λ replaces δ.

REMARK 22: We will utilize Remark 15 of the cell cost operator [1] for finding the dual variables u^+_i, v^+_j when the operator δE^{+}_{ef} is applied given that $(e, f) \epsilon C^*_k$ (cases (3) and (4) of (61)). We will define sets S_1 and S_2 as in Remark 15 so that all arguments given in Remark 15 hold here also. Hence we can give an algorithm similar to A9(b).

ALGORITHM A13(b). For finding the dual variables u_i and v_j when the operator δE^{+}_{ef} is applied given that $(e, f) \epsilon C^*_k$ (cases (3) and (4) of (61)).

(0) Denote C_k by C and C^*_k by C^*. Let $(i, j) = (e, f)$. If $p(i) = j$, set $\lambda = -\delta u_i/(e_{ij} + \delta)\lambda \geq 0)$, and $m_i = +1$. Let $u_i = u_i + m_i\lambda$, $t = d(i) = 1$, $SR = \{i\}$ and go to (1). If $p(j) = i$, set $\lambda = -\delta u_i$ $(\lambda \leq 0)$ and $n_j = +1$. Let $v_j = v_j + n_j\lambda$, $t = d(j) + 1$, $SC = \{j\}$ and go to (3).

Steps (1)–(5) are the same as those in A9(b) [1], where λ replaces δ.

REMARK 23: Due to similarities in Algorithms A9, [1] for the cost operator and A13 for weight operator, (as far as the dual variables are concerned), the maximum extent μ_λ^+ for λ can be given similar to (40), (41) of [1]. From this value of μ_λ^+, it is easy to compute the maximum dual extent μ_D^+ for the operator δE_{pq}^+. These are given below for cases (1)–(4) of (61).

$$(69)\quad \begin{cases} \mu_D^+ = \max \begin{cases} 0 \\ -\mu_\lambda^+ e_{pq}(\alpha-\beta)/(\beta u_p e_{pq} + \alpha \mu_\lambda^+) \end{cases} & \text{for case (1) of (61)} \\ \mu_D^+ = \max \begin{cases} 0 \\ -\mu_\lambda^+ e_{pq}(\alpha-\beta)/(\beta u_p e_{pq} + \beta \mu_\lambda^+) \end{cases} & \text{for case (2) of (61)} \\ \mu_D^+ = \max \begin{cases} 0 \\ -\mu_\lambda^+ e_{pq}/(u_p + \mu_\lambda^+) \end{cases} & \text{for case (3) of (61)} \\ \mu_D^+ = -\mu_\lambda^+/u_p & \text{for case (4) of (61)} \end{cases}$$

If the expressions given in the denominator for μ_D^+ in the above cases become zero, set $\mu_D^+ = M$ (a large positive number).

When the operator δE_{pq}^- is applied, the dual variables are obtained by putting $-m_i$ and $-n_j$ for m_i and n_j in algorithms A13(a) and A13(b), similar to the cost operator δC_{pq}^-. These results are summarized in Proposition 4.

PROPOSITION 4:

(a) For δE_{pq}^+ with $(p, q) \in B$ the new dual solutions are given by

$$(70)\quad u_i^+ = \begin{cases} u_i + m_i\lambda & \text{for } i \in I_p \\ u_i & \text{for } i \in I_q \end{cases}$$

$$v_j^+ = \begin{cases} v_j + n_j\lambda & \text{for } j \in J'_q \\ v_j & \text{for } j \in J'_p, \end{cases}$$

where m_i, n_j, and λ are functions of e_{ij}, u_i and δ as given in algorithms A13(a) and A13(b) and sets I_p, J'_q, I_q and J'_p are defined in Definition 17 [1].

The maximum extent μ_λ^+ for λ to which these new dual variables (70) are basis preserving if predecessor of p is q, is given by

$$(71)\quad \mu_\lambda^+ = \text{Min} \begin{cases} (e_{ij}u_i + v_j - c_{ij})/|n_j| & \text{for } (i,j) \in [(I_q \times J'_q) \cap UB] \\ (c_{ij} - e_{ij}u_i - v_j)/e_{ij}m_i & \text{for } (i,j) \in [(I_p \times J_p) \cap LB] \end{cases}$$

and if $p(q)=p$,

$$\mu_\lambda^+ = \text{Min} \begin{cases} (e_{ij}u_i + v_j - c_{ij})/e_{ij}|m_i| & \text{for } (i,j) \epsilon [(I_p \times J'_p) \cap UB] \\ (c_{ij} - e_{ij}u_i - v_j)/n_j & \text{for } (i,j) \epsilon [(I_q \times J'_q) \cap LB]. \end{cases} \tag{72}$$

The maximum dual extent μ_D^+ is

$$\mu_D^+ = f(\mu_\lambda^+), \tag{73}$$

where the functional expression is given in Equation (69) and μ_λ^+ as given by (71) or (72).

(b) For δE_{pq}^- with $(p, q) \epsilon B$, the new dual solutions are given by

$$\begin{aligned} u_i^- &= \begin{cases} u_i - m_i\lambda & \text{for } i \epsilon I_p \\ u_i & \text{for } i \epsilon I_q \end{cases} \\ v_j^- &= \begin{cases} v_j - n_j\lambda & \text{for } j \epsilon J'_q \\ v_j & \text{for } j \epsilon J'_p, \end{cases} \end{aligned} \tag{74}$$

where m_i, n_j and λ are the same as in part (a).

The maximum extent μ_λ^+ for λ to which these new dual variables (74) are basis preserving, if predecessor of p is q, is given by

$$\mu_{\bar{\lambda}} = \text{Min} \begin{cases} (c_{ij} - e_{ij}u_i - v_j)/|n_j| & \text{for } (i,j) \epsilon [(I_q \times J'_q) \cap LB] \\ (e_{ij}u_i + v_j - c_{ij})/e_{ij}m_i & \text{for } (i,j) \epsilon [(I_p \times J'_p) \cap UB] \end{cases} \tag{75}$$

and if $p(q) = \text{row } p$,

$$\mu_{\bar{\lambda}} = \begin{cases} (c_{ij} - e_{ij}u_i - v_j)/e_{ij}|m_i| & \text{for } (i,j) \epsilon [(I_q \times J_p) \cap LB] \\ (e_{ij}u_i + v_j - c_{ij})/n_j & \text{for } (i,j) \epsilon [(I_q \times J'_q) \cap UB] \end{cases} \tag{76}$$

The maximum dual extent $\mu_{\bar{D}}$ is

$$\mu_{\bar{D}} = f(\mu_{\bar{\lambda}}) \tag{77}$$

where the functional expression is given in Equation (69) and μ_λ^+ given by (75) or (76) replaces μ_λ^+ given in (69). (In (71) and (76) $J_p = J'_p \cup \{(n+1)\}$.)

PROOF: Proof is similar to that of Theorem 7 [1].

EXAMPLE (continued): Consider Figure 21, where δE_{22}^+ is applied where $(2,2) \epsilon B$. Since $p(\text{row } 2) =$ Col. 2 case (1) of (61) is applicable. Here $\alpha = 12$ and $\beta = 2$. Let us apply algorithm A13(a) to find the

$\lambda = \lambda_D = \delta/(10\delta + 25)$

u_i \ v_j →	$(31/5) - 2\lambda$	$(28/5) - \lambda$	3	4	0	a_i
0	3	6	(3)	(4)	(0)	300
	15	N	40	80	N	
$-1/5 + 2\lambda$	(6)	$3+\delta$ (5)	8	15	0	225
	150	75	25	N	N	
$-13/10 + \lambda/2$	(1)	(3)	10	6	0	140
	60	N	30	50	N	
0	100	100	100	100	(0)	1000
	N	N	N	N	N	
b_j	170	60	35	60	—	

FIGURE 21. Dual changes due to δE_{22}^{+}.

transformed dual variables and the maximum dual extent μ_D^+. The following gives the steps of A13(a) and their results.

Steps of A13(a)	*Result*
(0)	$(i, j) = (2, 2)$; $\lambda = -2\delta\ (-1/5)\ 3/[12\delta + 30] = \delta/5(2\delta + 5)$; $n_2 = -1$ and $v_2 = v_2 - \lambda$. Go to (1).
(1)	$R_3 = p(C_2)$; $m_3 = 1/2$; $u_3 = u_3 + 1/2\lambda$; $i = R_3$.
(2)	$C_1 = p(R_3)$; $n_1 = -4 \times 1/2 = -2$; $v_1 = v_1 - 2\lambda$. Let $j = C_1$. Go to (1).
(1)	$R_2 = p(C_1)$; $m_2 = +2/1$; $u_2 = u_2 + 2\lambda$. Since $R_2 = e$ go to (3).
(3)	$t = 1$; $C_1 = \phi$; STOP.

Thus from proposition 4,

$$v_2^+ = v_2 - \lambda$$

$$u_3^+ = u_3 + 1/2\lambda$$

$$v_1^+ = v_1 - 2\lambda$$

$$u_2^+ = u_2 + 2\lambda$$

where $\lambda = \delta/5(2\delta + 5)$.

Other dual variables are unchanged. μ_λ^+ from (71) is given by

$$\mu_\lambda^+ = \text{Min} \begin{cases} 16/10 \\ 6/10,\ 12/20,\ \textcircled{1/10},\ 48/5,\ 33/5,\ 13/5 = 1/10 \text{ at cell (2, 5) (entering cell).} \end{cases}$$

Since the denominator for μ_D^+ is zero, $\mu_D^+ = M$ as given in (69), where M is a large positive number.

EXAMPLE (continued): Since we got the primal variables for P^T when δE_{22}^+ was applied with $\delta = 5$ (Figure 21), let us see the dual changes the same δ. If $\delta = 5$, $\lambda = \delta/5(2\delta + 5) = 1/15$. Since this $\lambda = 1/15 < \mu_\lambda^+ = 1/10$ given above this operator is basis preserving. Utilizing the results given by A13(a), the dual variables D^+ for problem P^T are given below:

$$v_2^+ = v_2 - \lambda = 28/5 - 1/15 = 83/15$$

$$u_3^+ = u_3 + \lambda/2 = -13/10 + 1/30 = -10/15$$

$$v_1^+ = v_1 - 2\lambda = 31/5 - 2/15 = 91/15$$

$$u_2^+ = u_2 + 2\lambda = -1/5 + 2/15 = -1/15$$

Other dual variables are unchanged. These results are shown in Figure 22.

$\lambda = \lambda_D = \delta/(10\delta + 25) = 1/15$

u_i \ v_j	91/15	83/15	3	4	0	a_i
0	3 | 3 | 15	2 | 6 | N	2 | (3) | 40	3 | (4) | 80	1 | (0) | N	300
−1/15	1 | (6) | 150	8 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
−19/15	4 | (1) | 60	2 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 22. Dual changes due to $5E_{22}^+$.

The following theorem can now be stated and proved utilizing propositions 3 and 4.

THEOREM 16:

(a) For δE_{pq}^+ with $(p, q) \epsilon B$

x_{ij}^{+} is given by (64)
u_i^{+}, v_j^{+} are given by (70)
$Z^{+}-Z$ is given by (68)

and the maximum extent μ^{+} to which this basis preserving operator can be applied is given by

(78) $$\mu^{+}=\text{Min}\,\{\mu_P^{\ddagger}, \mu_D^{\ddagger}\}, \text{ where}$$

$\mu_P^{\ddagger}$ and $\mu_D^{\ddagger}$ are given by Equations (63) and (69), respectively.

(b) For $\delta E_{\bar{p}q}^{-}$ with $(p, q)\epsilon B$,

$x_{\bar{i}\bar{j}}$ is given by (65)
$u_{\bar{i}}$, $v_{\bar{j}}$ are given by (74)
$Z^{-}-Z$ is given by (68)

and the maximum extent μ^{-} to which this basis preserving operator can be applied is given by

(79) $$\mu^{-}=\text{Min}\,\{\mu_{\bar{P}}, \mu_{\bar{D}}\}, \text{ where}$$

$\mu_{\bar{P}}$, and $\mu_{\bar{D}}$ are given by Equations (67) and (77), respectively.

PROOF:

(a) Due to proposition 3, x_{ij}^{+} as given by (64) is primal feasible and due to proposition 4, u_i^{+} and v_j^{+} as given by (70) are dual feasible. Since $\mu^{+}\leqslant\mu_P^{\ddagger}$ as well as $\leqslant\mu_D^{\ddagger}$, for $0\leqslant\delta\leqslant\mu^{+}$, the transformed problem P^{T} is both primal and dual feasible and hence optimal.

(b) Proof of this part is similar.

EXAMPLE (continued): The primal variables are given in Figure 20 when δE_{22}^{+} is applied with $\delta=5$ and the dual variables are given in Figure 22 when the same operator δE_{22}^{+} with $\delta=5$ is applied. Since this $\delta=5\leqslant\mu^{+}=\text{Min}\,(\mu_P^{\ddagger}, \mu_D^{\ddagger})=M$ these changes in both primal and dual variables are primal and dual feasible respectively. Thus the solution is optimal from Theorem 16(a). These results are given in Figure 23. The total cost change from (68) is

$$Z^{+}-Z=\lambda_P\left[\sum_{(i,j)\epsilon\Gamma_1\cup\Gamma_2} c_{ij}m_{ij}\right]$$

$$=20[3\times 1+6\times 1/2-5\times 1-1\times 1/2]=10.$$

REMARK 24: It is necessary to compare the λ_P as given in algorithm A12(a) (b) with λ_D as given in algorithms A13(a), (b) when δE_{ef}^{+} is applied. Without any loss of generality let us consider δE_{pq}^{+} for case (1) of (61), where $p(e)=f$ and $(e, f)\epsilon C$. $\lambda_P=-\alpha\delta x_{ef}/[e_{ef}(\beta-\alpha)-\alpha\delta]$ as given in A12(a) while for the same case, $\lambda_D=\delta\beta u_e e_{ef}/[e_{ef}(\beta-\alpha)-\alpha\delta]$ as given in A13(a). When we compare these two expressions it is obvious that the denominator is the same while the numerator is negative in both cases (since we assume e_{ij}'s to be positive and $u_i\leqslant 0$ at the optimal). This causes the situation viz, when $\mu_P^{+}=M$ then $\mu_D^{+}=M$ and vice versa. These will be further discussed in a different paper.

u_i \ v_j	91/15	83/15	3	4	0	a_i
0	3 | 15 | 3 | 15	2 | 6 | N	2 | 35 | (3) | 40	3 | 60 | (4) | 80	1 | 5 | (0) | N	300
−1/15	1 | 145 | (6) | 150	8 | 10 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
−19/15	4 | 10 | (1) | 60	2 | 50 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 23. Primal and dual changes due to $5E^{+}_{22}$ (combination of Figs. 19 and 21).

ALGORITHM A14: For applying the basis preserving weight operator $\delta E^{\pm}_{pq}$ with $(p, q) \epsilon B$ and computing the maximal extent.

(1) Set the data of the problem $P^{\pm}$ same as problem P except that $e^{\pm}_{pq} = e_{pq} \pm \delta$. Set $B^{\pm} = B$, $LB^{\pm} = LB$, $UB^{\pm} = UB$. Find $\mu^{\pm}_{\lambda}$ for primal from (66). Find sets I_p, I_q, J'_P, J'_q. If predecessor of p is q, go to (2). Else go to (3).

(2) Find $\mu^{\pm}_{\lambda}$ for dual from (71), ((75)). Find $\mu^{\pm}_{P}$ from (67) and $\mu^{\pm}_{D}$ from (73), ((77)). Go to (4).

(3) Find $\mu^{\pm}_{\lambda}$ for dual from (72), ((76)). Find $\mu^{\pm}_{P}$ from (67) and $\mu^{\pm}_{D}$ from (73), ((77)).

(4) Compute $\mu^{\pm} = \text{Min}\ [\mu^{\pm}_{P}, \mu^{\pm}_{D}]$. Then for $0 \leq \delta \leq \mu^{\pm}$ find $\lambda^{\pm}$ for primal from A12 and $\lambda^{\pm}$ for dual from A13. With this value of (primal) λ find $X^{\pm}$ from (65) and $Z^{\pm} - Z$ from (68). Also find $u^{\pm}_{i}$ for $i \epsilon I$ and $v^{\pm}_{j}$ for $j \epsilon J'$ from (70), ((74)) utilizing (dual) λ. STOP.

BIBLIOGRAPHY

[1] Balachandran, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Generalized Transportation Problem—II; Cost, Rim and Bound Operators." Nav. Res. Log. Quart. 22, 101–125 (1975).

[2] Balachandran, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Generalized Transportation Problem—III; Weight Operators," Nav. Res. Log. Quart., 22, this issue, June (1975).

[3] Balachandran, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Generalized Transportation Problem—IV; Global Operators," Nav. Res. Log. Quart., 22, this issue, June (1975).

[4] Balachandran, V. and G. L. Thompson, "The Four Index Algorithm for the Generalized Transportation Problem," Management Science Research Report #310, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pa.

[5] Balas, E., "The Dual Method for the Generalized Transportation Problem," Management Science, Vol. 12, No. 7 (Mar. 1966).

[6] Balas, E. and P. L. Ivanescu, "On the Generalized Transportation Problem," Management Science, Vol. 11, No. 1 (Sept. 1964).

[7] Berge, C., *The Theory of Graphs and Its Applications* (John Wiley and Sons, Inc., New York, New York, 1962).

[8] Charnes, A. and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming* (John Wiley and Sons, New York, New York, 1961), Vols I and II.

[9] Dantzig, G. B., *Linear Programming and Extensions* (Princeton University Press, 1963).

[10] Eisemann, K., "The Generalized Stepping Stone Method for the Machine Loading Model," Management Science, Vol. 11, No. 1 (Sept. 1964).

[11] Eisemann, K. and J. R. Lourie, "The Machine Loading Problem," IBM 704 Program, IBML-1, IBM Applications Library, New York (1959).

[12] Ferguson, A. R. and G. B. Dantzig, "The Allocation of Aircraft to Routes—An Example of Linear Programming Under Uncertain Demand," Management Science, Vol. 3, No. 1 (1956).

[13] Garwin, W. W., *Introduction to Linear Programming* (McGraw-Hill Book Company, New York, New York, 1960).

[14] Glicksman, S., L. Johnson, and L. Eselson, "Coding the Transportation Problem," Nav. Res. Log. Quart., *7*, 169–183 (1960).

[15] Glover, F. and D. Klingman, "On the Equivalence of Some Generalized Network Problems to Pure Network Problems," Mathematical Programming, *4*, 269–278 (1973).

[16] Glover, F. and D. Klingman, "A Note on Computational Simplifications in Solving Generalized Transportation Problems," Transportation Science, *7*, 351–361 (1973).

[17] Hadley, G., *Linear Programming* (Addison-Wesley, Reading, Mass., 1962).

[18] Harary, F., *The Theory of Graphs* (Addison-Wesley, Reading, Mass., 1970).

[19] Lourie, J. R., "Topology and Computation of the Generalized Transportation Problem, *Management Science*, Vol. 11, No. 1 (Sept. 1964).

[20] Simmonard, M., *Linear Programming* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966).

[21] Srinivasan, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Transportation Problem—I," Nav. Res. Log. Quart., *19*, 205–225 (1972).

[22] Srinivasan, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Transportation Problem—II," Nav. Res. Log. Quart., *19*, 227–252 (1972).

[23] Srinivasan, V. and G. L. Thompson, "Accelerated Algorithms for Labelling and Relabelling of Trees, with Application to Distribution Problems," J. of the ACM, *19*, 712–726 (1972).

[24] Balachandran, V. and G. L. Thompson, "An Operator Theory for the Generalized Transportation Problem—I, Basic Theory," Nav. Res. Log. Quart. *22*, 79–100 (1975).

AN OPERATOR THEORY OF PARAMETRIC PROGRAMMING FOR THE GENERALIZED TRANSPORTATION PROBLEM– IV–GLOBAL OPERATORS*

V. Balachandran

Northwestern University
Evanston, Illinois

and

G. L. Thompson

Carnegie-Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

This paper investigates the effect of the optimal solution of a (capacitated) generalized transportation problem when the data of the problem (the rim conditions—i.e., the available time of machine types and demands of product types, the per unit production costs, the per unit production time and the upper bounds) are continuously varied as a linear function of a single parameter. Operators that effect the transformation of optimal solution associated with such data changes, are shown to be a product of basis preserving operators (described in our earlier papers) that operate on a sequence of adjacent basis structures. Algorithms are furnished for the three types of operators—rim, cost, and weight. The paper concludes with a discussion of the production and managerial interpretations of the operators and a comment on the "production paradox."

1. INTRODUCTION

In this paper we shall complete the development of an "Operator Theory" of parametric programming for the generalized transportation problem that was initiated in [24] and developed in [1] and [2]. The reader is assumed to be fully familiar with the three references cited above before attempting a study of this paper.

The operators that were studied in [1] and [2] were assumed to be basis preserving (Definition 15 in [24]), i.e., the original problem P and the transformed problem P^T share a common basis structure (basis B and sets UB and LB of variables at lower and upper bounds, respectively) in their (basic) optimal solutions. In section 2, we present a general framework for extending the operator theory to operators that are not necessarily basis preserving. It is shown that an operator $\delta^T(P)$ can be represented as a product of basis preserving operators that operate on a sequence of "adjacent" basis structures. The detailed algorithms for applying rim, cost and weight operators are then presented in sections 3, 4, and 5, respectively. In the final section, 6, we provide managerial interpretations of

*This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Contract N00014-67-A-0314-0007 NR 047-048 with the U.S. Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

the operators in determining shadow prices (i.e., the rate of change of optimal cost associated with a nonzero change in the data of the problem).

In this paper also, we shall use the same notation, bibliography, and examples as [24, 1, 2] and will continue from [2], the numbering of theorems, lemmas, remarks, equations, algorithms, figures, etc. Some of the properties are similar to those of [22] and hence proofs are not given though the properties are stated.

2. REPRESENTATION OF AN OPERATOR AS A PRODUCT OF BASIS PRESERVING OPERATORS

Our aim in parametric programming is to determine the optimum solution of a problem P^{T} (as defined in (22)) as a function of the non-negative parameter δ. By Assumption (A2) the system of constraints (6)–(8) is feasible for the problem P^{T} when $\delta=0$. Lemma 1 establishes a property regarding this feasibility as function of δ.

LEMMA 1: Consider the system of constraints (6)–(8) for problem P^{T} as defined in (22) [24]. If there is no feasible solution to P^{T} for a value of $\delta=\delta'$ then there is no feasible solution to P^{T} for $\delta>\delta'$.

PROOF: Assume the contrary—i.e., there exists a $\delta''>\delta'$ for which a feasible solution $\{x''_{ij}\}$ exists satisfying (6)–(8) to the problem P^{T} *with* $\delta=\delta''$. i.e.,

$$\sum_{j\in J} e_{ij}x''_{ij}=a_i+\delta''\alpha_i \qquad \text{for } i\in I, \tag{80}$$

$$\sum_{i\in I} x''_{ij}=b_j+\delta''\beta_j \qquad \text{for } j\in J', \text{ and} \tag{81}$$

$$0\leq x''_{ij}\leq U_{ij}+\delta''\nu_{ij} \qquad \text{for } i\in I', j\in J'. \tag{82}$$

For the problem P^{T} with $\delta=0$, $\{x_{ij}\}$ gives a feasible solution to (6)–(8) by Assumption (A2). Let us construct an intermediate solution, $\{x'_{ij}=(1-\delta'/\delta'')x_{ij}+(\delta'/\delta'')x''_{ij}\}$ and assert that it is feasible for P^{T} with $\delta=\delta'$. The rest of the proof is similar to that given for Lemma 1 of [22].

By Lemma 1 and the fact that P^{T} is feasible for $\delta=0$ due to Assumption (A2), it follows that P^{T} is feasible over a connected region $0\leq\delta\leq\delta'$. The fact that this region is closed can be shown by writing the constraints (6)–(8) for the problem P^{T} in the variables x_{ij} and δ as

$$\sum_{j\in J} e_{ij}x_{ij}-\delta\alpha_i=a_i \qquad \text{for } i\in I$$

$$\sum_{i\in I} x_{ij}-\delta\beta_j=b_j \qquad \text{for } j\in J'$$

$$x_{ij}\geq 0 \text{ and } x_{ij}-\delta\nu_{ij}\leq U_{ij} \qquad \text{for } i\in I, j\in J.$$

Since the above system of constraints defines a closed convex set [20] the region of feasible δ is closed as well.

Let us assume that the optimum solution of P^{T} is sought for δ in the range $0\leq\delta\leq\delta^*$. Then, as shown earlier, there exists a δ' $(0\leq\delta'\leq\delta^*)$ such that P^{T} is (primal) feasible only for the range

$0 \leq \delta \leq \delta'$. If $\delta' = \delta^*$, then the problem is feasible throughout the range of interest. The procedure to be presented shortly for applying $\delta^*\mathbf{T}(P)$ is such that δ' is identified so that the optimum solution for every δ in the range $0 \leq \delta \leq \delta'$ is found. Thus this method for applying $\delta^*\mathbf{T}(P)$ is basically a parametric programming procedure, but it could be used for post-optimization as well, i.e., to transform the optimum solution of P to that of a problem $\hat{P}$ that results when a discrete change is made in the data of the problem. (Note that we do not make any change in the weights e_{ij}, $(i, j) \in (I \times J)$.) For instance, if a_i, b_j, c_{ij} and U_{ij} are changed to $\hat{a}$, $\hat{b}$, $\hat{c}_{ij}$ and $\hat{U}_{ij}$ for the problem $\hat{P}$ (satisfying assumption (A1)), and if we define

$$\alpha_i = \hat{a}_i - a_i \qquad \text{for } i \in I'$$

$$\beta_j = \hat{b}_j - b_j \qquad \text{for } j \in J'$$

$$\left.\begin{aligned} \gamma_{ij} &= \hat{c}_{ij} - c_{ij} \\ \nu_{ij} &= \hat{U}_{ij} - U_{ij} \end{aligned}\right\} \qquad \text{for } (i, j) \in (I' \times J'),$$

the optimum solution for $\hat{P}$ can be obtained by applying $\delta^*\mathbf{T}(P)$ with $\delta^* = 1$ (see (22)). (If there is no feasible solution to $\hat{P}$, then the procedure will identify a number $\delta' \geq 0$ beyond which no primal feasible solution exists.)

Let us consider the application of an operator $\delta^*\mathbf{T}$ to the problem $P^1 = P$ with an optimal basis structure $B^1 = B$, $LB^1 = LB$, $UB^1 = UB$. (Superscript k will be used to denote the kth problem P^k.) Let μ^1 be the maximum extent to which the basis preserving operator δT can be applied to P^1. (Algorithm A7, A10, and A14 determine μ^T for a rim, cost or weight operator, respectively. In section 4 of [1] these results were extended to other operators also). If $\delta^* \leq \mu^1$ then we have the optimal solution for any $\delta (0 \leq \delta \leq \delta^*)$ by applying the *single* basis-reserving operator $\delta T(P^1)$. But, if $\delta^* > \mu^1$ this solution will *not* be optimal, (i.e., we will lose primal or dual feasibility, c.f. algorithm A8, if we applied $\delta R(P^1)$ with $\delta > \mu^1$, the leaving cell $(e_1, f_1) \in B^1$ would violate (8); if we apply $\delta C(P^1)$ with $\delta > \mu^1$ in A11, the entering cell $(e_1, f_1) \in LB^1$ (or UB^1) will violate dual feasibility condition 18 (or (19)); if we applied $\delta E(P^1)$ with $\delta > \mu^1$, the cell (e_1, f_1) (leaving or entering depending upon $\mu_P \leqq \mu_D$) would violate either primal feasibility conditions (8) or dual feasibility condition (18) or ((19)) depending on $\mu^1 = \mu_P$ or $\mu^1 = \mu_D$ from A14 of [2]. The same logic is applicable to any general $\delta T(P^1)$ since in section 4 of [1] is is shown that many basis-preserving operators can be reduced to the simultaneous application of a rim and a cost operator.

Let us denote by P^2 the problem obtained by applying $\mu^1 T$ to P^1 and by X^2, D^2 and $Z^2 = F^2$ the associated optimal solution. From Definition 14, it directly follows that applying $\delta\mathbf{T}(P^1)$ with $\delta > \mu^1$ is the same as applying $(\delta - \mu^1)\mathbf{T}(P^2)$. But as shown earlier the operator $T(P^1)$ is applicable only up to an extent of μ^1 with (B^1, LB^1, UB^1) so that $T(P^2)$ can be applied only over a zero range with (B^1, LB^1, UB^1). To apply $\mathbf{T}(P^2)$ over a positive range we follow the usual procedure of parametric programming of the linear program [20], and find an *adjacent basis structure* (B^2, LB^2, UB^2) different from (B^1, LB^1, UB^1) which is also optimal for T^2 (i.e., solution X^2 (or D^2) gets modified by this change, but Z^2 remains the same). In the case of rim operators a cell (e_1, f_1) was identified as the leaving cell (Algorithm A7) and in section 3, we provide the procedure for determining a nonbasic cell (g_2, h_2) so

that (B^2, LB^2, UB^2) is optimal for T_2 with $B^2 = B^1 - \{(e_1, f_1)\} + \{(g_2, h_2)\}$. (Of course, it is possible that there may not exist such a cell (g_2, h_2) which will be shown to imply that $\delta' = \mu^1$ gives the maximum range over which $\delta\mathbf{T}(P)$ can be applied,—i.e., the transformed problem P^T has no feasible solution for $\delta > \delta' = \mu^1$). In the case of cost operators a nonbasic cell (e_1, f_1) was identified as an *entering* cell (A10) and in section 4 we provide a procedure for finding a cell (g_2, h_2) such that (B^2, LB^2, UB^2) is also optimal for T^2 with $B^2 = B^1 + \{(e_1, f_1)\} - \{(g_2, h_2)\}$. We will show in section 4 that for cost operators it is possible that $B^2 = B^1$ and the change is only in the sets LB and UB. In the case of weight operator, a cell $(e_1, f_1) \epsilon B^1$ if (i) $\mu^1 = \mu_P$ (63) or a nonbasic cell, say, (e_1, f_1) if (ii) $\mu^1 = \mu_D$ (69) was identified as a *leaving cell* in case (i) or as an *entering cell* in case (ii) in Algorithm A14. In section 5, we will use the results of sections 3 and 4, and provide a procedure for determining a nonbasic cell (g_2, h_2) for case (i) or a cell (g_2, h_2) for case (ii) such that (B^2, LB^2, UB^2) is also optimal for T_2 with $B^2 = B^1 \mp \{(e_1, f_1)\} \pm \{(g_2, h_2)\}$. (The $\mp (e_1, f_1)$ and $\pm (g_2, h_2)$ will be according as case (i) or (ii), respectively.) In section 4 of [1], it was shown that any general basis preserving operator δT can be represented as $\delta T(P) = \delta R[\delta C(P)] = \delta C[\delta R(P)]$ and that $\mu(T) = \text{Min } [\mu(R), \mu(C)]$. The basis change for problem P_2 is to be applied based on the rim operator (cost operator) if $\mu(T) = \mu(R)$ $(\mu(T) = \mu(C))$. For the case of $\mu(T) = \mu(R) = \mu(C)$ either of the basis changes can be made.

Thus, we now find μ^2, the maximum extent to which the basis preserving operator can be applied to P^2 with (B^2, LB^2, UB^2). (Of course, it is possible that $\mu^2 = 0$; but this does not imply that we cannot apply $\delta\mathbf{T}(P^2)$ with $\delta > 0$, —that may be possible after obtaining some other adjacent basis structure). If $\delta^* \leq \mu^1 + \mu^2$, then we have the optimal solution, since we can apply $\delta T(P^1)$ for $0 \leq \delta \leq \mu^1$, and for $\mu^1 \leq \delta \leq \mu^1 + \mu^2$ we can first obtain P^2, X^2, D^2, and $Z^2 = F^2$ by applying $\mu^1 T(P^1)$, thereby transforming (B^1, LB^1, UB^1) to (B^2, LB^2, UB^2) and apply $(\delta - \mu^1)T(P^2)$, i.e., apply $(\delta - \mu^1)[\mu^1 T(P^1)]$. However, if $\delta^* > \mu^1 + \mu^2$, we will have to find P^3 (by applying $\mu^2 T$ to P^2) modify (B^2, LB^2, UB^2) to (B^3, LB^3, UB^3) (if possible) with associated changes in X^3 or D^3 or both and find μ^3. Then for $\mu^1 + \mu^2 < \delta \leq \mu^1 + \mu^2 + \mu^3$, the relevant operator is given by

$$(\delta - \mu^1 - \mu^2)T(P^3) = (\delta - \mu^1 - \mu^2)[\mu^2 T(P^2)] = (\delta - \mu^1 - \mu^2)[\mu^2 T(\mu^1 T(P))] = (\delta - \mu^1 - \mu^2) \prod_{k=1}^{2} \mu^k T(P).$$

Continuing in this manner we will prove that we obtain a finite sequence of "corner" [22] problems $\{P^k\}$ given by

$$P = P^1, P^2, \ldots, P^k, \ldots, P^\tau, \tag{83}$$

and an associated sequence of basis structures $\{(B^k, LB^k, UB^k)\}$:

$$(B, LB, UB) = (B^1, LB^1, UB^1), \ldots (B^k, LB^k, UB^k), \ldots (B^\tau, LB^\tau, UB^\tau) \tag{84}$$

where $(B^\tau, LB^\tau, UB^\tau)$ is optimal for the transformed problem P^T with $\delta = \delta'$ (as mentioned earlier P^T is feasible only for $0 \leq \delta \leq \delta'$ where $\delta' \leq \delta^*$—if $\delta' = \delta^*$ then $(B^\tau, LB^\tau, UB^\tau)$ is optimal for P^T with $\delta = \delta^*$), and such that any two consecutive basis structures $(B^{k-1}, LB^{k-1}, UB^{k-1})$ and (B^k, LB^k, UB^k) are different and optimal for $P^k (2 \leq k \leq \tau)$. Then for any $0 \leq \delta \leq \delta'$, the operator $\delta\mathbf{T}(P)$ is presented as

(85) $$\delta\mathbf{T}(P) = \lambda T(P^t) \text{ or equivalently}$$

(86) $$\delta\mathbf{T}(P) = \lambda T\left[\mu^{t-1}T(\mu^{t-2}T(\ldots(\mu^1 T(P))))\right]$$

$$= \lambda T\left[\prod_{k=1}^{t-1} \mu^k T(P)\right]$$

where $0 \leq t \leq \tau$ is an integer such that

(87) $$\sum_{k=1}^{t-1} \mu^k \leq \delta \leq \sum_{k=1}^{t} \mu^k \text{ and}$$

(88) $$\lambda = \delta - \sum_{k=1}^{t-1} \mu^k.$$

(The convention $\sum_{k=1}^{0} (\cdot) = 0$ and $\prod_{k=1}^{0} (\cdot) = 1$, is followed here also.)

The specific details for changing the basis structures and proof of finiteness are given in sections 3, 4, and 5 for rim, cost, and weight operators, respectively. For rim operators applying $\mu^k R(P^k)$ changes X^k to X^{k+1} but leaves D^k unaltered. (Theorems 4, 5). In section 3 the change from (B^k, LB^k, UB^k) to $(B^{k+1}, LB^{k+1}, UB^{k+1})$ is such that the primal solution X^{k+1} remains unaltered, but the dual solution gets changed from D^k to D^{k+1}. (X^{k+1} and D^{k+1} are the primal and dual solutions to T^{k+1} with $(B^{k+1}, LB^{k+1}, UB^{k+1})$). The reverse is true for cost operators while both are true for weight operators. Applying $\mu^k C(P^k)$ changes D^k to D^{k+1}, but leaves X^k the same (Theorem 9). But the change from (B^k, LB^k, UB^k) to $(B^{k+1}, LB^{k+1}, UB^{k+1})$ in section 4 is such that the dual solution D^{k+1} remains the same and X^k gets modified to X^{k+1}. Applying $\mu^k E(P^k)$ changes D^k to D^{k+1} and X^k to X^{k+1} (Theorem 16) and the change from (B^k, LB^k, UB^k) to $(B^{k+1}, LB^{k+1}, UB^{k+1})$ in section 5 is such that both primal solution X^{k+1} and dual solution D^{k+1} are modified.

3. PARAMETRIC PROGRAMMING WITH RIM OPERATORS

Since area and cell operators share most of the results that are common, we use δR to denote any rim operator. Let μ^k denote the maximum extent to which the rim operator δR can be applied to the problem P^k with basis B^k. Algorithm 7, in addition to providing μ^k, determines the leaving cell $(e_1, f_1) \in B^k$, $((e_1^k, f_1^k)$ to be precise, but for ease in later notation, we have dropped the superscripts and the subscripts in $(e_1^k, f_1^k))$ i.e. the cell that reaches either its lower or upper bound on applying $\mu^k R(P^k)$ and violates (8) if $\delta R(P^k)$ is applied with $\delta > \mu^k$, will be denoted by just (e, f) instead of (e_1^k, f_1^k). To determine the problem P^{k+1} and the optimal primal solution X^{k+1} with cost Z^{k+1}, we apply $\mu^k R(P^k)$ while using algorithm A8 (the dual matrix D_k is not affected by this transformation), so that $x_{e,f} = 0$ or $U_{e,f}$. (Both subscripts 1 and superscripts k are dropped.) We provide below how the entering cell (g, h) is determined so that the basis $B^{k+1} = B^k - \{(e, f)\} + \{(g, h)\}$ is dual feasible. This change in the basis alters the dual solution u_i^k and v_j^k to u_i^{k+1} and v_j^{k+1}.

Methods for finding adjacent dual feasible bases are given in literature, for instance in Balas

[5] and Charnes and Cooper [8]. We provide a method based on the method of Srinivasan and Thompson [22] for the ordinary transportation problem. This is based on the partitioning of sets I and J' similar to $I_p, I_{qk}, I_q, J'_q, J'_{pk}, J'_p$ as given in Definition 17 [1]. Following Definition 17 partition I and J' by the cell (e,f) i.e. we define the following sets given $(e,f)\epsilon B_l^k$

(89) $\Omega_l = B_l^k - \{(e,f)\}$ and set

(90)
$$\begin{cases} I_e = \{e\} \cup \{i \epsilon I \mid i \text{ is connected to } e \text{ in } \Omega_l\} \\ I_{fl} = I_l - I_e \\ I_f = I - I_e \\ J'_q = \{f\} \cup \{i \epsilon J' \mid j \text{ is connected to } f \text{ in } \Omega_l\} \\ J'_{el} = J'_l - J'_f \\ J'_e = J' - J'_f \end{cases}$$

(Note if the primes are removed in either of the three sets of columns then that set is the union of the corresponding set with prime and column $(n+1)$). If $(e,f)\epsilon C_l^k$ then $I_e = I_l$ and $J'_f = J'_l$ so that I_{fl} and J'_{el} are empty. An algorithm for finding these sets is given in [4].

We will now proceed to a lemma which requires a set of nonbasic cells. This set will be labelled as Ψ and consists of nonbasic cells as given below:

Let $\Psi = \Psi_1 \cup \Psi_2$ where

(91)
$$\Psi_1 = \begin{cases} (I_f \times J'_f) \\ (I_f \times J'_f); \\ (I_f \times J'_{el}); \end{cases} \qquad \Psi_2 = \begin{cases} (I_e \times J'_f) & \text{for } (e,f)\epsilon C_l^{k*} & \text{(branch)} \\ (I_e \times J_e) & \text{for } (e,f)\epsilon C_l^k & \text{(cycle)} \\ \{(I_e \times J'_e) - B^k\} & \text{if } f = n+1 & \text{(loop)} \end{cases}$$

With this set $\Psi = \Psi_1 \cup \Psi_2$ as shown above the following lemma is true.

LEMMA 2: Let Ψ be a set of nonbasic cells as given by (91). Then $B^k - \{(e,f)\} + \{(g,h)\}$ is a basis if and only if $(g,h)\epsilon\Psi$.

The proof of this lemma is based on the following facts.

(a) $(g,h)\epsilon\Psi \Rightarrow B^k - \{(e,f)\} + \{(g,h)\}$ is a basis.

(b) $(g,h)\notin\Psi$ and nonbasic $\Rightarrow B^k - \{(e,f)\} + \{(g,h)\}$ is either a two-tree or a component without a cycle or loop which are impossible due to the properties of the basis.

The details of the proof are similar to those of Lemma 2 of [22] and hence omitted.

EXAMPLE (continued): Let us consider the problem $P = P^1$ given in Figure 4 of [24] with basis B^1. The cell rim operator δR_{23}^+ was applied to this problem in [1] and the changes are given in Figure 4 in [1]. It was shown that $\mu^1 = 2.5$, since applying δR_{23}^+ with $\delta > 2.5$ would violate the non-negativit

ondition for the cell $(e^1, f^1) = (e, f) = (1, 5)$. The problem P^2 and the solution X^2 are shown in Figure 4 where $x^2_{ef} = x^2_{15} = 0$. However, the dual solutions u^1_i for $i \in I$ and v^1_j for $j \in J'$ are unaltered as given in igure 4 [24].

Let us now partition the sets I and J by $(e, f) = (1, 5)$ utilizing Equation (90) and obtain $I_e = \{1\}$; $_{fl} = \emptyset$; $I_f = \{2, 3, 4\}$, $J'_f = \{5\}$, $J'_{el} = \{3, 4\}$; $J'_e = \{1, 2, 3, 4\}$. Then according to Lemma 2, the set of *onbasic cells* Ψ are given such that if $(g, h) \in \Psi$, then $B^1 - \{(1, 5)\} + \{(g, h)\}$ is a basis, where $\Psi = \{(2, 3), (2, 4), (3, 3), (3, 4), (4, 3), (4, 4), (1, 1), (1, 2)\}$. The reader can verify that $B^1 - \{(1, 5)\} - \{(g, h)\}$, where $(g, h) \in \Psi$ does in fact, constitute a basis (Lemma 2) though the total number of ne-trees in B^1 may be reduced. The other nonbasic cells in Figure 24 are $\{(4, 1), (4, 2), (2, 5), (3, 5)\}$ ontained in $\overline{\Psi}$ the complement of Ψ from the nonbasic cells. It is easy to see that if $(g, h) \in \overline{\Psi}$, then $^1 - \{(1, 5)\} + \{(g, h)\}$ has a *two-tree* which does not constitute a basis thus verifying Lemma 2.

Theorem 17, given below, determines the bases (if any) that are dual feasible from among the djacent bases $B^k - \{(e, f)\} + \{(g, h)\}$ with $(g, h) \in \Psi$. Two cases are to be considered depending on hether $x^{k+1}_{ef} = 0$ or U_{ef}.

THEOREM 17: Consider a basis of kth stage B^k and the set of nonbasic cells LB^k and UB^k. Let $_e$, I_{fl}, I_f, J'_f, J'_{el} and J'_e be the sets as defined in (90) obtained due to partitioning of I and J' by the cell $e, f) \in B^k_l$. Find Δ by (i) or (ii) below depending upon whether $x^{k+1}_{ef} = 0$ or U_{ef}.

(i) $x^{k+1}_{ef} = 0$. If $p(e) = f$, let

$$\Delta = \text{Min} \begin{cases} (c_{ij} - e_{ij}u_i - v_j)/|n_j| & \text{for } (i, j) \in [\Psi_1 \cap LB^k] \\ (e_{ij}u_i + v_j - c_{ij})/e_{ij}m_i & \text{for } (i, j) \in [\Psi_2 \cap UB^k]. \end{cases} \tag{92}$$

f $p(f) = e$, let

$$\Delta = \text{Min} \begin{cases} (c_{ij} - e_{ij}u_i - v_j)/e_{ij}|m_i| & \text{for } (i, j) \in [\Psi'_2 \cap LB^k] \\ (e_{ij}u_i + v_j - c_{ij})/n_j & \text{for } (i, j) \in [\Psi_1 \cap UB^k]. \end{cases} \tag{93}$$

(ii) $x^{k+1}_{ef} = U_{ef}$. If $p(e) = f$, let

$$\Delta = -\text{Min} \begin{cases} (e_{ij}u_i + v_j - c_{ij})/|n_j| & \text{for } (i, j) \in [\Psi_1 \cap UB^k] \\ (c_{ij} - e_{ij}u_i - v_j)/e_{ij}m_i & \text{for } (i, j) \in [\Psi'_2 \cap LB^k] \end{cases} \tag{94}$$

ınd, if $p(f) = e$

$$\Delta = -\text{Min} \begin{cases} (e_{ij}u_i + v_j - c_{ij})/e_{ij}|m_i| & \text{for } (i, j) \in [\Psi_2 \cap UB^k] \\ (c_{ij} - e_{ij}u_i - v_j)/n_j & \text{for } (i, j) \in [\Psi_1 \cap LB^k], \end{cases} \tag{95}$$

vhere Ψ_1 and Ψ_2 are given by (91) for different cases and Ψ'_2 obtained from putting J' instead of $'$ in Ψ_2 as defined in (91).

(a) If the minimum of (92), (93), (94), or (95) is attained at a cell (g, h) then $B^{k+1}=B^k-\{(e,f)\}+\{(g, h)\}$ is an adjacent dual feasible basis with sets LB^{k+1}, UB^{k+1} defined accordingly and the duals u_i^{k+1} for $i\epsilon I$ and v_j^{k+1} for $j\epsilon J'$ are given by (96) below:

$$u_i^{k+1}=\begin{cases} u_i^k-m_i\Delta & \text{for } i\epsilon I_e \\ u_i & \text{for } i\epsilon I_f \end{cases}$$

(96)

$$v_j^{k+1}=\begin{cases} v_j^k-n_j\Delta & \text{for } j\epsilon J'_f \\ v_j & \text{for } j\epsilon J'_e, \end{cases}$$

where m_i, n_j are functions of e_{ij} as given by Algorithms A9(a) or A9(b) of [1] at the kth stage.

(b) If the set over which the minimization is done is empty, then there is no primal feasible solution to the problem obtained by applying the operator $\delta\mathbf{R}(P^{k+1})$ with $\delta>0$.

PROOF: The proof of this theorem is similar to that given for Theorem 11 of [22] and the proof of Theorem 7 given in our earlier paper [1]. In these m_i and n_j are obtained via algorithms A9(a) or A9(b) of [1] for the basis structure and duals of the kth stage. For ease of notation we did not put the superscript k for m_i, n_j. The details of the proof are omitted.

The next two lemmas determine the effect of the basis change from B^k to B^{k+1} on the cost effects of the cell and area rim operators—i.e., on the expressions $\{d_{pq}\}$ and $\left(\sum_{i\epsilon I'} e_{ij}\alpha_i u_i+\sum_{j\epsilon J'}\beta_j v_j\right)$. (See Theorem 3 [1]).

LEMMA 3:

(a) For the operator δR_{pq}^+, $(p, q)\epsilon\Psi_1$, when $p(e)=f$ and $\epsilon\Psi_2$ when $p(f)=e$, if $x_{ef}^{k+1}=0$; and $(p, q)\epsilon\Psi_2$ when $p(e)=f$ and $\epsilon\Psi_1$ when $p(f)=e$, if $x_{ef}^{k+1}=U_{ef}$.

(b) For the operator δR_{pq}^-, $(p, q)\epsilon\Psi_2$ when $p(e)=f$ and $\epsilon\Psi_1$, when $p(f)=e$, if $x_{ef}^{k+1}=0$; and $(p, q)\epsilon\Psi_1$ when $p(e)=f$ and $\epsilon\Psi_2$ when $p(f)=e$ if $x_{ef}^{k+1}=U_{ef}$.

(c) For the operator δR_A,

$$\sum_{i\epsilon I_e} e_{ij}\alpha_i-\sum_{j\epsilon J'_e}\beta_j<0 \quad \text{if } x_{ef}^{k+1}=0 \quad \text{and} \quad \sum_{i\epsilon I_e} e_{ij}\alpha_1-\sum_{j\epsilon J'_e}\beta_j>0$$

if $x_{ef}^{k+1}=U_{rs}$, where Ψ_1 and Ψ_2 are as given by Equation (91).

PROOF: The proof of this lemma is similar to that of Lemma 3 of [22] and proof of Theorem 7 of our earlier paper [1] and the properties of the two-tree given in [24]. Thus the proof is omitted.

LEMMA 4:

(a) For the operator δR_{pq}^+, $d_{pq}^{k+1}\geq d_{pq}^k$

(b) For the operator δR_{pq}^-, $d_{pq}^{k+1}\leq d_{pq}^k$

(c) For the operator δR_A,

$$\sum_{i\epsilon I} e_{ij}\alpha_i u_i^{k+1}+\sum_{j\epsilon J'}\beta_j v_j^{k+1}\geq\sum_{i\epsilon I} e_{ij}\alpha_i u_i^k+\sum_{j\epsilon J'}\beta_j v_j^k.$$

Proof of this lemma is similar to that given for Lemma 4 of [22] and hence omitted.

We now summarize the results of this rim global operator section by the following algorithm.

ALGORITHM A15: For applying the operator $\delta^*\mathbf{R}(P)$.

(1) Let $k=1$, $P^1=P$, $B^1=B$, $LB^1=LB$, $UB^1=UB$, $X^1=X$, $Z^1=Z=F$, $u_i^1=u_i$ for $i\epsilon I$ and $v_j^1=v_j$ for $j\epsilon J'$.

(2) Utilizing algorithm A7 determining μ^k and the leaving cell (e,f). If $\sum_{i=1}^{k}\mu^j \geq \delta^*$, set $\delta'=\delta^*$ and $t'=k$ and go to (4). Otherwise apply the operator $\mu^k R(P^k)$ using algorithm A8. Define the resulting problem P^{k+1} with optimal primal solution X^{k+1} and cost Z^{k+1} (the dual variables u_i^k and v_j^k are not altered due to this transformation).

(3) Utilizing Equation (90) partition the sets I and J' by the cell (e,f) into $I_e, I_{\bar{f}}, I_f, J_f', J_{\bar{e}}'$ and J_e'. Use (92), (93), (94), or (95) (depending upon $p(e)=f$ or $p(f)=e$ and $x_{ef}^{k+1}=0$ or U_{ef} as given in Theorem 17) to determine Δ and the cell (g,h) (if any). If no such cell (g,h) exists (i.e., the set over which the minimization is done in Theorem 17 is empty) set $\delta'=\sum_{i=1}^{k}\mu^i$ and $t'=k$ and go to (4). Otherwise set $B^{k+1}=B^k-\{(e,f)\}+\{(g,h)\}$. Obtain the sets LB^{k+1} and UB^{k+1} accordingly. Determine the new duals u_i^{k+1} for $i\epsilon I$ and v_j^{k+1} for $j\epsilon J'$ utilizing (96). Set $k=k+1$ and go to (2).

(4) The problem P^T defined by the operator $\delta\mathbf{R}(P)$ is primal feasible over the range $0\leq\delta\leq\delta'$. To find the optimum solution for any δ in this range determine t and λ as defined by (87) and (88) and apply $\lambda R(P^t)$ using Algorithm A8. STOP.

REMARK 25: An expression for the optimal cost Z as a function of δ in the range $0\leq\delta\leq\delta'$ will ow be provided. By (85)–(88) and algorithm A8 [1], it follows that

7) $$Z^+(\delta)=Z^1+\sum_{k=1}^{t-1}\mu^k d_{pq}^k+\lambda d_{pq}^t \qquad \text{for the operator } \delta\mathbf{R}_{pq}^+,$$

8) $$Z(\delta)=Z^1-\sum_{k=1}^{t-1}\mu^k d_{pq}^k-\lambda d_{pq}^t \qquad \text{for the operator } \delta\mathbf{R}_{pq}^-, \text{ and}$$

99) $$Z^A(\delta)=Z^1+\left[\sum_{k=1}^{t-1}\mu^k\left[\sum_{i\epsilon I}e_{ij}\alpha_i u_i^k+\sum_{j\epsilon J'}\beta_j v_j^k\right]\right]+\lambda\left[\sum_{i\epsilon I}e_{ij}\alpha_j u^t{}_i+\sum_{j\epsilon J'}\beta_j v^t{}_j\right] \qquad \text{for the operator}$$

$\mathbf{R}^A$, where t and λ are as given by (87) and (88), respectively.

Thus it is easy to see that the cost curves are piecewise linear. From Lemma 4 and Equations 7)–(99), the cost function $Z(\delta)$ is seen to be convex, thus conforming with the well known result in arametric programming (of the b vector) for the general linear programming problems.

EXAMPLE (continued): Let us continue in Figures 24–27 the application of the same rim operator R_{23}^+ discussed earlier in Figures 7 of [1] and in Figure 24, in this paper. Let us assume that we are nterested in the optimum solution of the transformed problem for all values of δ, $0\leq\delta\leq\delta^*=5$ (say). Utilizing Algorithm A15 where $P_1=P$ is the problem of Figure 4 [24] with optimum solution X^1, basis ^{1}, duals u_i^1 and v_i^1, and cost $Z^1=F^1=1{,}460$. It was shown in section 2 of [1], that application of Algorithm A7 for δR_{23}^+ can be done to a maximum of $\mu^1=2.5$ with $(e,f)=(1,5)$ as the leaving cell. (The

optimum solution for $0 \leq \delta \leq 2.5$ is shown in Figure 7 with a cost effect of $d^1_{23} = 2.$) In Step (2) of Algorithm A15 we find $\mu^1 = 2.5 < \delta^* = 5$ so that we apply $2.5R^+_{23}(P^1)$ (using Algorithm A8) to obtain the problem P^2 shown in Figure 24 with optimal primal solution X^2 and cost $Z^2 = Z^1 + \mu^1 d^1_{23} = 1{,}460 + 2.5(2) = 1{,}465$.

In Step (3) we partition I and J by $(e, f) = (1, 5)$ to get $I_e = \{1\}$; $I_{fl} = \emptyset$; $I_f = \{2, 3, 4\}$; $J'_f = \{5\}$; $J'_{el} = \{3, 4\}$ and $J'_e = \{1, 2, 3, 4\}$. Since $x^2_{15} = 0$, we use the minimization (92), to obtain (g, h) and Δ. (Note that (1, 5) is a loop and by definition, row 1 is the root and arc (1, 5) is connected to node 1 itself.) Since the cell (1, 5) is slack cell we need a new algorithm (A9(c)) to find m_i and n_j.

It may be necessary to find the change in dual variables (section 3 of [1]) when the operator $\delta C^{\pm}_{e,n+1}$ is applied. We will given an algorithm for $\delta C^+_{e,n+1}$ and the dual variables for $\delta C^-_{e,n+1}$ are obtained. by replacing m_i by $-m_i$ and n_j by $-n_j$.

ALGORITHM A9(c): For finding the dual variables u_i and v_j when the operator $\delta C^+_{e,n+1}$ is applied given that $(e, n+1) \epsilon C_k \subseteq B_k$.

(0) Since there is no dual variable for $n+1$th column and the original u_e was 0, let $m_e = 1$ so that $u_e = u + m_e\delta = \delta$. Let $t = 1$ and $i = e$ and $SR = \{i\}$.
C^*_k.) (1)–(5) same as steps (1)–(5) of algorithm A9(b) given in [1].

(1) (Evaluation of u_i, v_j for cells in C^*_k.) (1)–(5) same as steps (1)–(5) of algorithm A9(b) given in [1].)

Utilizing this algorithm, it is easy to see that $m_1 = 1$, $n_3 = -2$, $n_4 = -3$and all other m_i $(i = 2, 3, 4)$ and n_j $(j = 1, 2)$ are 0 when $\delta C^+_{1,5}$ is applied to the optimal solution given in Figure 24. Thus minimizing over the candidate cells Ψ, $[(2, 3), (2, 4), (3, 3), (3, 4), (4, 3), (4, 4)] \cup [(1, 1), (1, 2)]$, we have $\Delta = \min[(8+1-3)/2,\ (15+2-4)/3,\ (10+(13/5)-3)/2,\ (6+(13/10)-4)/3,\ M,\ M,\ ((31/5)-3)/3] = 16/15$ for cell (1, 1). Thus, the entering cell $(g, h) = (1, 1)$ with $\Delta = 16/15$. Consequently we obtain the adjacent basis $B^2 = B^1 - \{(1, 5)\} + \{(1, 1)\}$ as shown in Figure 25 with $UB^2 = UB^1 - \{(1, 1)\} = \emptyset$

$P^2 = (5/2)R^+_{23}(P^1)$; (X^2, B^1, u^1_i, v^1_j)

u_i \ v_j →	31/5	28/5	3	4	—	a_i
0	3 | 15 | 3 | 15	2 | 6 | N	2 | 37.5 | (3) | 40	3 | 60 | (4) | 80	1 | 0 | (0) | N	300
−1/5	1 | 132.5 | (6) | 150	3 | 35 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	237.5
−13/10	4 | 22.5 | (1) | 60	2 | 25 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	37.5	60	—	

FIGURE 24

X^2, B^2, u_i^2, v_j^2 for $R^+_{23}(P)$

u_i \ v_j →	31/5	28/5	77/15	36/5	—	a_i
−16/15	3 15 (3) 15	2 6 N	2 37.5 (3) 40	3 60 (4) 80	1 0 N	300
−1/5	1 132.5 (6) 150	3 35 (5) 75	5 8 25	10 15 N	1 0 N	237.5
−13/10	4 22.5 (1) 60	2 25 (3) N	2 10 30	1 6 50	1 0 N	140
0	1 100 N	1 100 N	1 100 N	1 100 N	1 1000 (0) N	1000
b_j	170	60	37.5	60	—	

FIGURE 25

nd $LB^2 = LB^1 + \{(1, 5)\}$. The new duals u_i^2 and v_j^2 obtained from (96) are shown on left and top rims f the tableau given in Figure 25. The reader may verify that the primal solution and cost remain naltered due to this basis change and that the duals given in Figure 25 are dual feasible with this new asis structure (B^2, LB^2, UB^2). At this stage there are only two one-trees in the one-forest thus reduced y one from the three one-trees of the previous stage. Lemma 3(a) holds in this case since $(p, q) = (2, 3) \epsilon \Psi_1 = (I_f x J'_{et})$. Also $d_{23}^2 = e_{23}u_2^2 + v_3^2 = 5\ (-1/5) + 77/15 = 62/15 > d_{23}^1 = 2$ thus verifying Lemma 3(a). Let us now go back to Step (2) of Algorithm A15 with $k = 2$. In applying A7 to find μ^2 corresponding o $\delta R_{32}^+(P^2)$, we get the two tree Γ as shown in Figure 26 so that $\mu^2 = 2.5$ with $(e, f) = (1, 3)$ as the eaving cell ($m_{13} = +1$; $m_{11} = -2/3$; $m_{21} = -1/5$; $m_{22} = 26/15$; $m_{32} = -26/15$; $m_{31} = 13/15$). The optimum solution for $2.5 \leq \delta \leq 5 = \mu^1 + \mu^2$ is shown in Figure 26, where $\lambda = \delta - \mu^1 = \delta - 2.5$ (see (87) and (88)). Since $\mu^1 + \mu^2 = 5 = \delta^*$ we go to (4) of A15. The optimal solution for $\delta^* = 5$ is also the same solution as in Figure 27 and the optimal cost $Z^+ = 1{,}485.67$. In fact, for this problem $Z^+ = 1{,}460 + 2\delta$ for $0 \leq \delta \leq 2.5$; $1{,}465 + 62\delta/15$ for $2.5 \leq \delta \leq 5$; $1{,}485.67 + d_{23}^3\ \delta$ for $5 \leq \delta \leq \mu_3$ and so on. The values of d_{23}^3 and μ^3 can be calculated if $\delta^* > 5$. It is easily seen that these cost functions are piecewise linear and convex (since the slope increases) thus satisfying Remark 25 for this case.

To prove that Algorithm A15 terminates in a finite number of steps one needs an assumption hat the problem P is *dual nondegenerate.*

DEFINITION 19: The problem P defined by (5)–(8) [24], is *dual nondegenerate*, if corresponding o every dual feasible basis B with associated duals u_i for $i \epsilon I$ and v_j for $j \epsilon J'$ satisfying (17), conditions 18) and (19) are satisfied as strict inequalities.

REMARK 26: Perturbation techniques are available to preserve dual nondegeneracy [10, 20].

THEOREM 18: Assume that P is dual nondegenerate. Then Algorithm A15 terminates in a finite number of steps.

PROOF: Proof of this theorem is due to the fact Δ defined in Theorem 17 is nonzero and from Lemma 4. This is similar to the proof of Theorem 12 of [22] and thus omitted.

X for $2.5 \leq \delta \leq 5$; $\lambda = \delta - 2.5$

u_i \ v_j →	31/5	28/5	77/15	36/5	—	a_i
−16/15	3 | 15−2λ | (3) | 3 | 15	2 | 6 | N	2 | 37.5+λ | (3) | 40	3 | 4 | 80	1 | 4 | N	300
−1/5	1 | 132.5−λ | (6) | 150	3 | 35+26λ | (5) | 15 | 75	5 | (8) | 25	10 | 15 | N	1 | 0 | N	225
−13/10	4 | 22.5+13λ | (1) | 15 | 60	2 | 25−26λ | (3) | 15 | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 26

X^2, B^2, u_i^2, v_j^2 for $\delta R_{23}^+(P)$

u_i \ v_j →	31/5	28/5	77/15	36/5	—	a_i
−16/15	3 | 40/3 | (3) | 15	2 | 6 | N	2 | 40 | (3) | 40	3 | 60 | (4) | 80	1 | 0 | N	300
−1/5	1 | 132 | (6+δ) | 150	3 | 118/3 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	250
−13/10	4 | 74/3 | (1) | 60	2 | 62/3 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	40	60	—	

FIGURE 27

4. PARAMETRIC PROGRAMMING WITH COST OPERATORS

Assumption A2 in [24] shows that the initial problem is primal feasible. It is known that the cost operators affect only cost entries. Hence, it is clear that the transformed problem due to application of cost operator is also primal feasible for the entire range $0 \leq \delta \leq \delta^*$. We will use δC to denote the cost operator whether it is cell or area cost operator since the results for both of these are similar. Let μ^k be the maximum extent to which the basis preserving cost operator δC can be applied to P^k with basis structure (B^k, LB^k, UB^k). Algorithm A10 determines μ^k and the entering cell (e_k, f_k).

(It was e_1, f_1) in Algorithm A10 [1].) For convenience let us drop the subscript k in (e_k, f_k) and call it as just (e, f). This amount μ^k is such that the dual feasibility condition (18) or (19) will be violated by the entering cell (e, f) when the basis-preserving operator $\delta C(P^k)$ is applied with $\delta > \mu^k$. Let P^{k+1} be the problem (with optimum dual solution D^{k+1} and cost Z^{k+1}) resulting due to the application of $\mu^k C$ to P^k using Algorithm A11. (Note that the primal solution X^* remains unaltered due to this transformation.) Definition of μ^k and the choice of cell (e, f) leads to the following:

$$e_{ef}u_e^{k+1} + v_f^{k+1} = c_{ef}^{k+1}. \tag{100}$$

Thus the cell (e, f) satisfies (17)–(19) (though it is nonbasic) and hence qualifies as a basic cell or for a change from LB^k to UB^{k+1} or from UB^k to LB^{k+1}. How a cell (g, h) belonging to the current basis B^k is chosen in such a way that $B^{k+1} = B^k + \{(e, f)\} - \{(g, h)\}$ is also an optimal basis for P^{k+1} is described below. Alternately the only change in moving from (B^k, LB^k, UB^k) to $(B^{k+1}, LB^{k+1}, UB^{k+1})$ may be that (e, f) changes from LB^k to UB^{k+1} or from UB^k to LB^{k+1}. Since (e, f) already satisfies (17) the duals u_i^{k+1} for $i \epsilon I$ and v_j^{k+1} for $j \epsilon J'$ remain the same so that $(B^{k+1}, LB^{k+1}, UB^{k+1})$ is dual feasible. Consequently, the basis structure change is to be done in such a way that the primal solution X^{k+1} corresponding to $(B^{k+1}, LB^{k+1}, UB^{k+1})$ is primal feasible.

REMARK 27: Inclusion of a nonbasic cell (e, f) to an existing basis B^k, obtaining a two-tree Γ (refer to Algorithm A4 [24]) is a standard step in primal algorithms for the (capacitated) transportation problem [6, 10, 24] and hence no proof is needed here for the well known result that $B^{k+1} = B^k - \{(g, h)\} + \{(e, f)\}$ will be a basis (one-forest [24]) if and only if $(g, h) \epsilon \Gamma$. Since (e, f) satisfies (100) the change from X^k to X^{k+1} corresponds to finding an alternate optimum [24]. If $(g, h) \equiv (e, f)$ the new basis $B^{k+1} = B^k - \{(g, h)\} + \{(e, f)\}$ followed by appropriate changes in LB^k to LB^{k+1} and UB^k to UB^{k+1}. But, if $(g, h) = (e, f)$ the new basis $B^{k+1} = B^k$. The changes are only in the movement of (e, f) from LB^k to UB^{k+1} (if $x_{ef}^k = 0$) or from UB^k to LB^{k+1} (if $x_{ef}^k = U_{ef}$).

Theorem 19 below determines the cell $(g, h) \epsilon \Gamma$ (the two-tree of Equation (10) [24]) from among the adjacent bases $B^{k+1} = B^k - \{(g, h)\} + \{(e, f)\}$, such that the solution X^{k+1} associated with $(B^{k+1}, LB^{k+1}, UB^{k+1})$ is primal feasible. To do this let Γ be the two tree obtained as the union of row and column segments due to inclusion of (e, f) into the basis B^k so that the two sets Γ_1 and Γ_2 (where μm_{ij} are increased and decreased), respectively, from x_{ij} as given by Algorithm A4 [24] are defined.

THEOREM 19: Consider the solution X^k and the two sets Γ_1 and Γ_2 defined by Algorithm A) [24]) above. Find Δ by (i) or (ii) below depending on whether $x_{ef}^k = 0$ or U_{ef}.

(i) $x_{ef}^k = 0$. Let

$$\Delta = \text{Min} \begin{cases} x_{ij}^k/(-m_{ij}) & \text{for } (i, j) \epsilon \Gamma_2 \\ (U_{ij} - x_{ij}^k)/m_{ij} & \text{for } (i, j) \epsilon \Gamma_1. \end{cases} \tag{101}$$

(ii) $x_{ef}^k = U_{ef}^k$. Let

$$\Delta = -\text{Min} \begin{cases} (U_{ij} - x_{ij}^k)/m_{ij} & \text{for } (i, j) \epsilon \Gamma_2 \\ (x_{ij}^k)/-m_{ij} & \text{for } (i, j) \epsilon \Gamma_1. \end{cases} \tag{102}$$

Then the solution given below by (103) is primal feasible.

$$x_{ij}^{k+1} = \begin{cases} x_{ij}^k - m_{ij}\Delta & \text{for } (i,j)\epsilon\Gamma_2 \\ x_{ij}^k + m_{ij}\Delta & \text{for } (i,j)\epsilon\Gamma_1 \\ x_{ij}^k & \text{elsewhere,} \end{cases} \tag{103}$$

(a) If the minimum of (101) or (102) is attained at a cell $(g, h) \neq (e, f)$ then $B^{k+1} = B^k - \{(g,h)\} + \{(e,f)\}$ gives the adjacent basis with LB^k and UB^k appropriately modified to LB^{k+1} and UB^{k+1}.

(b) If the minimum of (101) or (102) is attained at (e, f) then the adjacent basis structure is given by $B^{k+1} = B^k$, with cell (e, f) moving from LB^k to UB^{k+1} (if $x_{ef}^k = 0$) or from UB^k to LB^{k+1} (if $x_{ef}^k = U_{ef}$).

PROOF: The proof of this theorem is essentially from Algorithm A4, which ensures primal feasibility due to coefficients m_{ij} for $(i, j)\epsilon\Gamma$ and from the way Equations (101), (102), and (103) are defined. Moreover a proof similar to that for Theorem 13 of [22] can be very easily given and hence is omitted.

The next two lemmas examine the effect of the change in the basis structures from stage k to stage $k+1$ and the associated change in the cost.

LEMMA 5:

(a) For the operator $\delta\mathbf{C}_{pq}^+$, $(p, q)\epsilon\Gamma_2$ if $x_{ef}^k = 0$ and $(p, q)\epsilon\Gamma_1$ if $x_{ef}^k = U_{ef}$.

(b) For the operator $\delta\mathbf{C}_{pq}^-$, $(p, q)\epsilon\Gamma_1$ if $x_{ef}^k = 0$ and $(p, q)\epsilon\Gamma_2$ if $x_{ef}^k = U_{ef}$.

(c) For the operator $\delta\mathbf{C}^A$, the amount $\sum_{(i,j)\epsilon\Gamma_1} \gamma_{ij} - \sum_{(i,u)\epsilon\Gamma_2} \gamma_{ij}$, is negative if $x_{ef}^k = 0$ and is positive if $x_{ef}^k = U_{ef}$.

PROOF: Proof of this lemma is similar to that of Lemma 5 of [22].

LEMMA 6:

(a) For the operator $\delta\mathbf{C}_{pq}^+$, $x_{pq}^{k+1} \leqslant x_{pq}^k$

(b) For the operator $\delta\mathbf{C}_{pq}^-$, $x_{pq}^{k+1} \geqslant x_{pq}^k$.

(c) For the operator $\delta\mathbf{C}_A$,

$$\sum_{i\epsilon I'} \sum_{j\epsilon J'} \gamma_{ij} x_{ij}^{k+1} \leqslant \sum_{i\epsilon I'} \sum_{j\epsilon J'} \gamma_{ij} x_{ij}^k .$$

PROOF: The proof is from Lemma 5 and due to the sign of Δ. Again, the proof of this lemma is similar to that of Lemma 6 of [22] and hence omitted.

The main results of this section are summarized and given in the following algorithm.

ALGORITHM A16: For applying the operator $\delta^*\mathbf{C}(P)$.

(1) Let $k=1$, $P^1 = P$; $B^1 = B$; $LB^1 = LB$; $UB^1 = UB$; $X^1 = X$; $Z^1 = Z$; $u_i^1 = u_i$ for $i\epsilon I$; and $v_j^1 = v_j$ for $j\epsilon J'$.

(2) By utilizing Algorithm A10, determine μ^k and the *entering cell* (e, f). If $\sum_{i=1}^{k} \mu_i \geqslant \delta^*$, set $\delta^1 = \delta^*$, $t' = k$ and go to (4). Else, apply $\mu^k C(P^k)$ utilizing Algorithm A11. Define the resulting problem as P^{k+1} with optimal dual solutions u_i^{k+1} for $i\epsilon I$ and v_j^{k+1} for $j\epsilon J'$ and cost Z^{k+1}. (Note that $X^{k+1} = X^k$ at this step.)

(3) Form the two-tree Γ obtained by including (e, f) to B^k and thus find the two sets Γ_1 and Γ_2 and m_{ij} for $(i, j) \epsilon \Gamma$ utilizing Algorithm A4 [24]. Use (101) or (102) depending upon the case $x^k_{ef}=0$ or $x^k_{ef}=U_{ef}$ and compute Δ. If the minimum is taken on a cell $(g, h) \epsilon B^k$, then define $B^{k+1}=B^k-\{(g, h)\}+\{(e, f)\}$ and adjust LB^k and UB^k to get LB^{k+1} and UB^{k+1}, respectively. If the minimum is taken on the cell (e, f) itself, move (e, f) from LB^k to UB^{k+1} if $x^k_{ef}=0$ or move (e, f) from UB^k to LB^{k+1} if $x^k_{ef}=U_{ef}$ and set $B^{k+1}=B^k$. Obtain X^{k+1} utilizing (103). Let $k=k+1$ and go to (2).

(4) To obtain optimal solution for a problem P^T defined by the operator $\delta \mathbf{C}(P)$, $(0 \leq \delta \leq \delta^*)$, obtain t and λ from (87) and (88) and then apply $\lambda C(P_t)$ utilizing Algorithm A11. STOP.

REMARK 28: An expression for the optimal cost Z as a function of δ in the range $0 \leq \delta \leq \delta^*$ will now be derived. By (85)–(88) and Algorithm A11, it follows that

$$Z^+(\delta)=Z^1+\sum_{k=1}^{t-1} \mu^k x^k_{pq}+\lambda x^t_{pq}, \qquad \text{for the operator } \delta \mathbf{C}^+_{pq},$$

$$Z^-(\delta)=Z^1-\sum_{k=1}^{t-1} \mu^k x^k_{pq}-\lambda x^t_{pq} \qquad \text{for the operator } \delta \mathbf{C}^-_{pq}, \text{ and}$$

$$Z^A(\delta)=A^1+\left[\sum_{k=1}^{t-1} \mu^k \left(\sum_{i \epsilon I'} \sum_{j \epsilon J'} \gamma_{ij} x^k_{ij}\right)\right]+\lambda \left(\sum_{i \epsilon I'} \sum_{j \epsilon J'} \gamma_{ij} x^t{}_{ij}\right) \qquad \text{for the operator } \delta \mathbf{C}^A,$$

where t and λ are defined by (87) and (88). Thus from the above expressions it is easy to see that the cost curves are piecewise linear. Confirming the well-known result in parametric programming of the 'price-vector,' it is seen from Lemma 6 and from the above equations, the cost functions $Z(\delta)$ form a concave function.

EXAMPLE (continued): We now illustrate with Figures 28–32, the application of the same cost operator $\delta \mathbf{C}^+_{pq}(P)$ discussed in section 3 of [1] with $(p, q)=(2, 1)$. Figure 4 [24] shows the problem $P^1=P$ with optimal solutions X^1, u^1_i for $i \epsilon I$ and v^1_j for $j \epsilon J'$ with basis structure (BJ^1, LB^1, UB^1) and optimal cost $Z^1=1{,}460$. In section 3 of [1] it was shown that the basis preserving operator $\delta \mathbf{C}^+_{21}(P^1)$ can be applied to a maximum extent of 2/3 with $(e_1, f_1)=(1, 2)$ as the entering cell. Assume that we are interested in the optimum solution for $0 \leq \delta \leq \delta^*=1{,}000$ (say). Then at step (2) of Algorithm A16, we find $\sum_{i=1}^{1} \mu^i < \delta^*$ so that $(2/3)C^+_{21}(P^1)$ is applied to obtain problem P^2. The optimal dual solutions u^2_i and v^2_j are shown in Figure 12 of [1] with cost $Z^2=Z^1+\mu^1 x^1{}_{21}=1{,}460+(2/3)(135)=1{,}550$. It is to be noted that the cell $(e, f)=(1, 2)$ satisfies (100). The optimal dual solution for $0 \leq \delta \leq 2/3$ is shown in Figure 11 of [1] with a cost of $Z=1{,}460+135\delta$. It is to be noted that the primal solution X^1 remains unchanged during this operation.

The two-tree Γ is formed in step (3) of Algorithm A16 by adding $(e, f)=(1, 2)$ to the basis B^1. This two-tree is shown in Figure 28 with the set $\Gamma_1=\{(1, 2), (3, 1), (2, 2)\}$ and $\Gamma_2=\{(3, 2), (2, 1), (1, 5)\}$. Its graph is shown in Figure 29. We utilize Algorithm A4 of [24] to get the multipliers m_{ij} for all $(i, j) \epsilon \Gamma$. They are given below: $m_{12}=1$; $m_{32}=-6/5$; $m_{31}=3/5$; $m_{21}=-3/5$; $m_{22}=1/5$ and $m_{15}=-2$. Since

X^1, B^1, u_i^2, v_j^2 for $C_{21}^+(P)$

u_i \ v_j →	7	6	3	4	—	a_i
0	3 | 15 | 3 | 15	2 | 6 | N | 15	2 | 35 | 3 | 40	3 | 60 | 4 | 80	1 | 5 | 0 | N	300
−1/3	1 | 135 | 6+δ | 150	3 | 30 | 5 | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
−3/2	4 | 20 | 1 | 60	2 | 30 | 3 | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | 0 | N	1000
b_j	170	60	35	60	—	

FIGURE 28

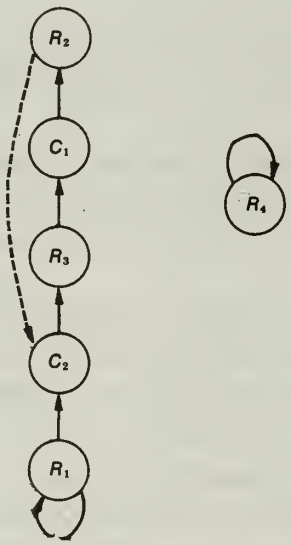


FIGURE 29. Graph associated with Figure 28.

$x_{ef}^1 = 0$, we apply (101) so that $\Delta = \min\{25, 225, 2.5, M, 66.6, 225\} = 2.5$ with $(g, h) = (1, 5)$. We now transform the solution X^1 as per (103) to obtain the solution X^2 shown in Figure 30, with $B^2 = B^1 - \{(1, 5)\} + \{(1, 2)\}$. The reader may verify that this solution is primal feasible, i.e. satisfies (6)–(8). Also $LB^2 = LB^1 + \{(1, 5)\} - \{(1, 2)\}$. Note that Lemmas 5(a) and 6(a) hold true. The cost effect has dropped from $x_{21}^1 = 135$ to $x_{21}^2 = 133.5$.

Returning back to step (2) of Algorithm A16, we determine μ^2 associated with the basis in Figure 30 using Algorithm A10. In step (3) of A10, we find $I_p = \{1, 2, 3\}$, $I_{q1} = \emptyset$; $I_q = 4$; $J'_q = \{1, 2, 3, 4\}$; $J'_{p1} = 0$, $J_p = \{5\}$. Following Algorithm A9(a), using the graph as given by Fig. 31 corresponding to Figure 30,

X^2, B^2, u_i^2, v_j^2 for $C_{21}^+(P)$

u_i \ v_j →	7	6	3	4	—	a_i
0	3 | 15 | 3 | 15	2 | 2.5 | (6) | N	2 | 35 | (3) | 40	3 | 60 | (4) | 80	1 | 0 | (0) | N	300
−1/3	1 | 133.5 | (6) | 150	3 | 30.5 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
−3/2	4 | 21.5 | (1) | 60	2 | 27 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 30

we have $u_2^2 = -1/3 - 1/5\lambda$; $v_2^2 = 6 + (3\lambda/5)$; $u_3^2 = -3/2 - (3\lambda/10)$; $v_1^2 = 7 + (6\lambda/5)$; $u_1^2 = 0 - 3/10\lambda$; $v_3^2 = 3 + 3\lambda/5$; and $v_4^2 = 4 + 9\lambda/10$. From (41), μ^2 is evaluated. But, it is seen that the set $[(I_p x J'_p) \cap UB]$ is empty while $[I_q x J'_q \cap LB]$ consists of the cells (4, 1) (4, 2) (4, 3) (4, 4) which corresponds to the fictitious row. The minimum value is $(M-7)/(6/5) = M'$ attained at cell (4, 1). Thus for the entire range $2/3 < \delta < M'$ this solution is optimal and for our $\delta^* = 1{,}000$ this is true. In other words whatever may be the cost increase beyond 2/3 at cell (2, 1) the solution remains optimal and thus insensitive to any cost variation. See Figure 32.

It is necessary to assume that the problem is primal nondegenerate to prove convergence of Algorithm A16.

DEFINITION 20: The problem P is *primal nondegenerate* if for any basis structure (B, LB, UB) the associated primal solution X (satisfying (6) and (7)) satisfies the following:

$$(i, j) \epsilon B \Rightarrow (0 < x_{ij} < U_{ij}). \tag{104}$$

REMARK 29: Perturbation techniques [8, 20] are available if it is found that the problem is primal degenerate, thus preserving primal nondegeneracy.

THEOREM 20: Algorithm A16 terminates in a finite number of steps assuming that the problem P is primal nondegenerate.

PROOF: The value of Δ obtained in (101) or (102) is nonzero due to Assumption (104). Thus the inequalities in the statements of Lemma 6(a)–(c) hold as strict inequalities. Thus the value of x_{pq} is strictly decreasing (increasing) for the operator $\delta C_{pq}^+ (\delta C_{pq}^-)$. So also, $\sum_{I \epsilon I'} \sum_{j \epsilon J'} \gamma_{ij} x_{ij}$ is strictly decreasing for the operator δC_A. These values are unique for a given basis structure (B, LB, UB) so that repetition of a basis structure is impossible. Since the total number of basis structures is finite algorithm A6 terminates in a finite number of steps.

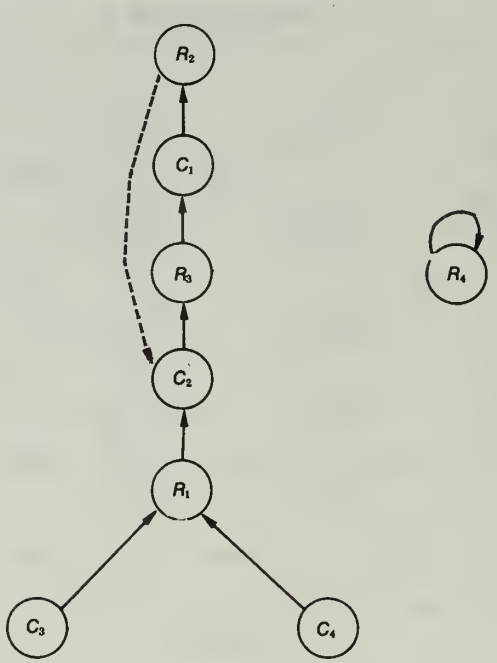


FIGURE 31. Graph associated with Figure 30.

u_i and v_j for $2/3 \leq \delta \leq \delta^*$ with $\lambda = \delta - 2/3$ for $C_{21}^+(P)$

u_i \ v_j →	$7+3\lambda/5$	$6+3\lambda/5$	$3+3\lambda/5$	$4+9\lambda/10$	—	a_i
$-3\lambda/10$	3 | 15 | 3 | 15	2 | 2.5 | (6) | N	2 | 35 | (3) | 40	3 | 60 | (4) | 80	1 | 0 | N	300
$-1/3$ $-\lambda/5$	1 | 133.5 | $(6+\lambda)$ | 150	3 | 30.5 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
$-3/2$ $-3\lambda/10$	4 | 21.5 | (1) | 60	2 | 27 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 32

5. PARAMETRIC PROGRAMMING WITH WEIGHT OPERATORS

In [2], the basis preserving weight operator $\delta E_{pq}^{\pm}$ was discussed. Let $\mu^k = \min\{\mu_p^k, \mu_n^k\}$ denote the maximum extent to which the weight operator δE (we have dropped subscript pq for convenience) can be applied to the problem P^k with basis B^k. Here μ_P^k and μ_D^k are the "maximum primal extent" and "maximum dual extent" as given in [2] associated with the problem P^k with basis B^k. Algorithm

A14 [2], in addition to providing μ^k, determines the *leaving cell* $(e^k, f^k) \epsilon B^k$ if $\mu^k = \mu_P^k$ or the *entering cell* (e^k, f^k) if $\mu^k = \mu_D^k$. Following sections 3 and 4 let us drop for convenience the superscript k in (e^k, f^k). Thus, if (e, f) is the *leaving cell* when $\mu^k = \mu_P^k$ then section 3 provided a method of finding an *entering cell* (g, h) so that the new basis $B^{k+1} = B^k - \{(e, f)\} + \{(g, h)\}$ is dual feasible. However, if (e, f) is the *entering cell* when $\mu^k = \mu_D^k$ then section 4 provided a method of finding a cell (*leaving* or changing the basis structure) (g, h) belonging to the current basis B^k is chosen such that $B^{k+1} = B^k + \{(e, f)\} - \{(g, h)\}$ is an optimal basis for P^{k+1}. (Alternately we also discussed the case of (e, f) changing from LB^k to UB^{k+1} or from UB^k to LB^{k+1}). Thus results similar to those of Lemmas 2, 3, and 4, Theorems 17 and 18, and Algorithm A15 are applicable to the weight operator δE when $\mu^k = \mu_P^k$ and results similar to those of Lemmas 5 and 6, Theorems 19 and 20 and Algorithm A16 are applicable to the weight operator δE when $\mu^k = \mu_D^k$. In all these Lemmas and Theorems referred above, we have to substitute the operator $\delta \mathbf{E}^+$ or δE^- for $\delta \mathbf{R}^+$ or $\delta \mathbf{R}^-$ when $\mu^k = \mu_P^k$ and when $\mu^k = \mu_D^k$, $\delta \mathbf{E}^+$ or δE^- replaces $\delta \mathbf{C}^+$ or $\delta \mathbf{C}^-$. The weight area operator is not discussed here and thus section (c) of the Lemmas and Theorems given earlier are not applicable to the case of weight operator. We need both primal and dual nondegeneracy assumption (Definitions 20 and 19) for proving the convergence of the algorithm for the weight operator. Thus we will give here the Algorithm A17 for applying the weight operator $\delta^* \mathbf{E}(P)$ utilizing the results given in both sections 3 and 4.

ALGORITHM A17: For applying the operator $\delta^* \mathbf{E}(P)$

(1) Let $k=1$; $P^1 = P$, $B^1 = B$, $LB^1 = LB$, $UB^1 = UB$, $X^1 = X$, $Z^1 = Z = F$, $u_i^1 = u_i$ for $i \epsilon I$, and $v_j^1 = v_j$ for $j \epsilon J'$.

(2) By utilizing Algorithm A14 [2], determine $\mu^k = \min(\mu_P^k, \mu_D^k)$ and the cell (e, f) which will be leaving if $\mu^k = \mu_P^k$ or entering the basis if $\mu^k = \mu_D^k$. If $\sum_{i=1}^{k} \mu^i \geq \delta^*$, set $\delta' = \delta^*$ and $t' = k$ and go to (4). Otherwise apply the operator $\mu^k E(P^k)$ by using Algorithm A14. Define the resulting problem P^{k+1} with

(a) Optimal primal solution X^{k+1} and cost Z^{k+1} (from Algorithm A15) if $\mu^k = \mu_P^k$ or

(b) Optimal dual solution u_i^{k+1} for $i \epsilon I$ and v_j^{k+1} for $j \epsilon J'$ (from Algorithm A16) if $\mu^k = \mu_D^k$.

(3) (a) If $\mu^k = \mu_P^k$, then apply step (3) of Algorithm A15.

(b) If $\mu^k = \mu_D^k$, then apply step (3) of Algorithm A16.

(4) (a) If $\mu^k = \mu_P^k$, then apply step (4) of Algorithm A15.

(b) If $\mu^k = \mu_D^k$, then apply step (4) of Algorithm A16.

EXAMPLE (continued): Let us now illustrate with Figures 33–35, the application of the same weight operator $\delta \mathbf{E}_{pq}^+(P)$ discussed in [2] with $(p, q) = (2, 2)$. Figure 4 [24] shows the problem $P_1 = P$ with optimal solutions X^1, u_i^1 for $i \epsilon I$ and v_j^1 for $j \epsilon J'$ with basis structure (B^1, LB^1, UB^1) and optimal cost $Z^1 = 1{,}460$. Let us assume, that we are now interested in the optimum solution for $0 \leq \delta \leq \delta^* = 1{,}000$ (say). But, in step (2) of Algorithm A17, when we determine $\mu^1 = \min(\mu_P^1, \mu_D^1)$ we know that μ_P^1 and μ_D^1 are large positive numbers say M_1 and M_2, respectively. Let us recall from [2] that $\mu_{\lambda P}^+ = 30$ where $\lambda_P = 60\delta/(5 + 2\delta)$ which gave $\mu_P^+ = (5\mu_{\lambda P}^+)/(60 - 2\mu_{\lambda P}^+) = M_1$ (since denominator is zero with $\mu_{\lambda P}^+ = 30$) and *leaving cell* is either (21) or (22). Further $\mu_{\lambda D}^+ = 1/10$ where $\lambda_D = \delta/(10\delta + 25)$ from [2] which gave $\mu_D^+ = 25\mu_{\lambda D}^+/(1 - 10\mu_{\lambda D}^+) = M_2$ (since denominator is zero with $\mu_{\lambda D}^+ = 1/10$) and the *entering cell* is (2, 5). Since $\delta^* < \min[M_1, M_2]$ we will apply the operator $\delta \mathbf{E}_{22}^1(P)$ for $0 \leq \delta \leq \delta^*$. When $\delta = 5$ ($\lambda_P = 20$ *and* $\lambda_D = 1/15$) we have the primal and dual changes given in Figure 23 of [2] with optimal cost

$\lambda_P = 60\delta/(5+2\delta)$; $\lambda_D = \delta/(10\delta+25)$

u_i \ v_j →	$31/5-2\lambda_D$	$28/5-\lambda_D$	3	4	0	a_i
0	3 | 15 | 3 | 15	2 | 6 | N	2 | 35 | (3) | 40	3 | 60 | (4) | 80	1 | 5 | (0) | N	300
$-1/5$ $+2\lambda_D$	1 | $135+(\lambda_P/2)$ | (6) | 150	3 | $30-\lambda_P$ | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
$-13/10$ $-\lambda_D/2$	4 | $20-(\lambda_P/2)$ | (1) | 60	2 | $30+\lambda_P$ | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 33. Primal and dual changes due to $\delta \mathbf{E}_{22}^{+}$ for $0 < \delta < M$

u_i \ v_j →	6.0001	5.5001	3	4	0	a_i
0	3 | 15 | 3 | 15	2 | 6 | N	2 | 35 | (3) | 40	3 | 60 | (4) | 80	1 | 5 | (0) | N	300
−0.0005	1 | 149.9999 | (6) | 150	1003 | 0.0001 | (5) | 75	5 | 8 | 25	10 | 15 | N	1 | 0 | N	225
−1.2501	4 | 15.0001 | (1) | 60	2 | 59.9999 | (3) | N	2 | 10 | 30	1 | 6 | 50	1 | 0 | N	140
0	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 100 | N	1 | 1000 | (0) | N	1000
b_j	170	60	35	60	—	

FIGURE 34. Primal and dual changes due to 1,000 E_{22}^{+}

1470. The primal and dual changes for $0 \leq \delta \leq 1{,}000$ are given in Fig. 33, when $\delta = \delta^* = 1{,}000$, $\lambda_P \simeq 29.9999$; and $\lambda_D \simeq 1/10.00001$ (Note that $\mu_{\lambda P}^{+} = 20$ and $\mu_{\lambda D}^{+} = 1/10$). Thus $x_{22} = 0.0001$; $x_{32} = 59.999$; $x_{31} = 15.00005$ and $x_{21} = 149.99905$. The cost Z^{+} from (68) of [2] will be $1{,}460 + \lambda_P/2 = 1474.99905$. It is to be noted that cost increases at the rate of $\lambda_P/2$ where $1/2 = \sum_{(i,j)\epsilon\Gamma_1 \cup \Gamma_2} c_{ij} m_{ij}^{+}$ from (68) of [2].

It is linear in λ_P, but λ_P is *nonlinear* in δ so that the cost curve is nonlinear in δ. The cost curve for $0 \leq \delta \leq \delta^* = 1{,}000$ is given in Figure 35. Now, turning our attention on the duals, when $\delta = 1{,}000$,

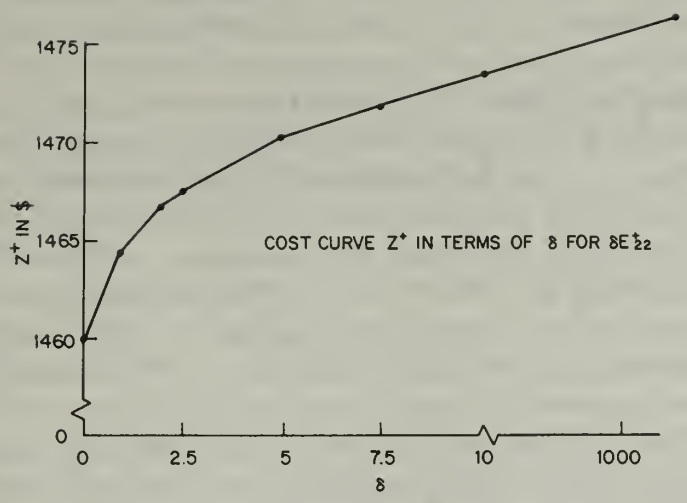


FIGURE 35

$\lambda_D \simeq 1/10.00001$ so that the duals when evaluated from (70) of [2] will be as follows

$$u_2 = (-1/5) + 2\lambda_D; \qquad u_3 = -13/10 + \lambda_D/2; \qquad v_1 = (31/5) - 2\lambda_D; \qquad \text{and} \qquad v_2 = (28/5) - \lambda_D.$$

The other duals are unchanged. The primal and dual changes due to $1{,}000\mathbf{E}_{22}^{+}$ are given in Figure 34. It is to be noted that in this case the basis preserving operator is global and the total optimal cost is bounded by 1,475.

6. PRODUCTION AND MANAGERIAL INTERPRETATIONS OF OPERATORS

In this final section of this paper, the effects on the optimal cost of increase (or decrease) in the capacity of available machine time a_i or the demand of product type b_j will be determined. A manager can utilize such information in making decisions as to which machine type capacities are to be increased, which product types should be produced more etc. Further, examining the effects of an increase (or decrease) in the unit costs of production (c_{ij}) and the decrease (or increase) of per unit production time of product j in machine i (e_{ij}) (more efficient production is associated with lesser value of e_{ij}) will help the manager in pricing and scheduling of products to the machine types. It can be seen that though the formulation of the problem (5)–(8) is based on production costs alone, by defining c_{ij} to be the negative of unit profit of manufacturing product j in machine type i, the machine loading problem (the generalized transportation model) (5)–(8) becomes a more general profit maximization problem.

Due to Theorem 3 [1], it follows that $d_{ij} = e_{ij}u_i + v_j$ gives the imputed value, i.e., the rate at which the optimal cost changes if the capacity of machine type i and the demand of product j were simultaneously changed. As shown in Figure 6 [24] d_{ij} could be negative though c_{ij}'s are all positive. Similarly by Theorem [1] $x_{ij}\,(-x_{ij})$ gives the shadow price when the unit cost c_{ij} is increased (decreased). By (8), $x_{ij} \geq 0$.

In the rim operator δR_{pq} both a_p and b_p were simultaneously changed. Suppose the manager is interested only in the increase of the capacity of machine type p, i.e., a_p is to be increased to $a_p + \delta$. Since the slack column $n+1$ has no constraint (b_{n+1} is not defined) and $e_{p,n+1} \equiv 1$ for all p this is a

particular case of $\delta\mathbf{R}_{p,n+1}$. Thus all the theorems and algorithms associated with "rim operators" [1] are applicable here. Moreover $d_{p,n+1}$ (the imputed value associated) $= e_{p,n+1}u_p$ which is ≤ 0 (all c_{ij}'s are assumed to be nonnegative, in the machine loading model) so that the optimal cost does not increase and by choosing the machine type p which has the most negative u_p maximum reduction in the total cost can be achieved. If a specific machine capacity p is increased the imputed value will be $u_p \leq 0$, (will be zero if $(p, n+1)$ is in the optimal basis). In this sense this u_p can be viewed as "downward marginal cost"—the rate at which the optimal cost can go down by a unit increase of available machine time per machine type p. Similarly, if the manager is interested in knowing the effect of increase in demand of product q, say from b_q to $b_q+\delta$, then we can apply the rim operator $\delta\mathbf{R}_{m+1,q}$. This is permissible, since the $m+1$th row, which corresponds to a fictitious machine has an unlimited capacity a_{m+1} and thus the shadow price $d_{m+1,q} = e_{m+1,q}u_{m+1} + v_q = v_q$ (note that $u_{m+1} \equiv 0$). Again since $c_{m+1,q} = M$ (a large positive quantity), $v_q \geq 0$ so that the optimal cost increases at the rate of v_q. Again choosing the product type associated with smallest v_q the total cost can be increased to the minimum limit. If a specific product, q, has a higher demand then the imputed cost associated with that is v_q. Note that u_p, v_q are all unique here unlike the transportation problem so that these shadow prices are real.

As in the ordinary transportation problem [22, p. 249] there is a "paradox" when $d_{pq} < 0$. For instance in Figure 6 of [24] suppose we make $e_{33} = 3$. Then, using the data in that figure we calculate $d_{33} = (-13/10)3 + 3 = -9/10$. Using the above interpretation of d_{pq} we see that if we apply δR^{+}_{33} we can reduce the total cost of production while increasing the total production, i.e., satisfying greater total demand. We can dubb this a "production paradox" to contrast it with the "transportation paradox" discussed in [22].

Further, if a specific production cost c_{pq} is increased (decreased) the corresponding change in cost is x_{pq} times the increase (decrease) and it is to be noted that this shadow price $x_{pq} \geq 0$. Similarly, when the machine utilization time e_{pq}, per unit of product q in machine p is increased (decreased) to $e_{pq}+\delta$ (or $e_{pq}-\delta$) when $(p, q) \in B$, the corresponding imputed value will be $\frac{\lambda}{\delta}\left[\sum_{(i,j)\in\Gamma_1 \cup \Gamma_2} e_{ij}m^{+}_{ij}\right]$. Though the imputed value is linear in λ, it should be noticed that λ is a nonlinear function of δ, except in the case when predecessor of $q = p$ and $(p, q) \in C^{*}_{k}$. (Refer to Equations (61), (68) and Proposition 3 of [2]). However, when (p, q) is not in the basis, the imputed value will be zero, when $(p, q) \in LB$ (Theorem 13), or when $(p, q) \in UB$ and $(p, n+1) \in B$ (Theorem 14). But when $(p, q) \in UB$ and $(p, n+1) \notin B$ the imputed value withh be $\mp U_{pq}u_p$ (Theorem 15). Further management applications and interpretation will be given in a different paper.

BIBLIOGRAPHY

[1] Balachandran, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Generalized Transportation Problem—II; Cost, Rim and Bound Operators." Nav. Res. Log. Quart. *22*, 101–125 (1975..

[2] Balachandran, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Generalized Transportation Problem—III; Weight Operators," Nav. Res. Log. Quart. *22*, this issue, June (1975).

[3] Balachandran, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the

Generalized Transportation Problem—IV; Global Operators," Nav. Res. Log. Quart., *22*, this issue, June (1975).

[4] Balachandran, V. and G. L. Thompson, "The Four Index Algorithm for the Generalized Transportation Problem," Management Science Research Report #310, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pa.

[5] Balas, E., "The Dual Method for the Generalized Transportation Problem," Management Science, Vol. 12, No. 7 (Mar. 1966).

[6] Balas, E. and P. L. Ivanescu, "On the Generalized Transportation Problem," Management Science, Vol. 11, No. 1 (Sept. 1964).

[7] Berge, C., *The Theory of Graphs and Its Applications* (John Wiley and Sons, Inc., New York, New York, 1962).

[8] Charnes, A. and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming* (John Wiley and Sons, New York, New York, 1961), Vols. I and II.

[9] Dantzig, G. B., *Linear Programming and Extensions* (Princeton University Press, 1963).

[10] Eisemann, K., "The Generalized Stepping Stone Method for the Machine Loading Model," Management Science, Vol. 11, No. 1 (Sept. 1964).

[11] Eisemann, K. and J. R. Lourie, "The Machine Loading Problem," IBM 704 Program, IBML–1, IBM Applications Library, New York (1959).

[12] Ferguson, A. R. and G. B. Dantzig, "The Allocation of Aircraft to Routes6an Example of Linear Programming Under Uncertain Demand," Management Science, Vol. 3, No. 1 (1956).

[13] Garwin, W. W., *Introduction to Linear Programming* (McGraw-Hill Book Company, New York, New York, 1960).

[14] Glicksman, S., L. Johnson, and L. Eselson, "Coding the Transportation Problem," Nav. Res. Log. Quart., *7*, 169–183 (1960).

[15] Glover, F. and D. Klingman, "On the Equivalence of Some Generalized Network Problems to Pure Network Problems," Mathematical Programming, *4*, 269–278 (1973).

[16] Glover, F. and D. Klingman, "A Note on Computational Simplifications in Solving Generalized Transportation Problems," Transportation Science, *7*, 351–361 (1973).

[17] Hadley, G., *Linear Programming* (Addison-Wesley, Reading, Mass., 1962).

[18] Harary, F., *The Theory of Graphs* (Addison-Wesley, Reading, Mass., 1970).

[19] Lourie, J. R., "Topology and Computation of the Generalized Transportation Problem, *Management Science*, Vol. 11, No. 1 (Sept. 1964).

[20] Simmonard, M., *Linear Programming* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966).

[21] Srinivasan, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Transportation Problem—I," Nav. Res. Log. Quart., *19*, 205–225 (1972).

[22] Srinivasan, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Transportation Problem—II," Nav. Res. Log. Quart., *19*, 227–252 (1972).

[23] Srinivasan, V. and G. L. Thompson, "Accelerated Algorithms for Labelling and Relabelling of Trees, with Application to Distribution Problems," J. of the ACM, *19*, 712–726 (1972).

[24] Balachandran, V. and G. L. Thompson, "An Operator Theory for the Generalized Transportation Problem—I, Basic Theory," Nav. Res. Log. Quart. *22*, 79–100 (1975).

A RECURSIVE BRANCH AND BOUND ALGORITHM FOR THE MULTIDIMENSIONAL KNAPSACK PROBLEM*

Arne Thesen

University of Wisconsin—Madison
Madison, Wisconsin

ABSTRACT

This paper presents an efficient branch and bound algorithm for the solution of certain multiconstrained knapsack problems. The key to this algorithm is a rigidly defined tree structure in which branching and bounding may be performed through recursive relationships. The algorithm is particularly useful when only limited amounts of core storage are available as only the current and one previous solution is saved at any one time. Execution speeds compare favorably with other algorithms. A numerical example and computational experience is given.

I. INTRODUCTION

The problem of choosing the best set of projects from a list of available alternatives subject to limited availability of resources can be abstracted into a fairly general problem frequently encountered in many areas.

When only one resource constraint is considered, the problem becomes the familiar knapsack or cargo loading problem, a problem which has received extensive attention in the literature. Kolesar [11,12] discusses several interesting applications of knapsack problems in the areas of project selection, capital budgeting, reliability, and cargo loading. Feuerman and Weiss [3] reports on an application in the area of test construction and scoring. Similar but multiconstrained problems are reported by Peterson in the project selection area [15] and Thesen in a computer program scheduling application [18].

Several approaches to the solution of knapsack problems have been suggested. Prior to the advent of efficient integer programming algorithms Bellman and Dreyfus offered a dynamic programming formulation [2], while Fulkerson presented a network approach [4]. Fairly efficient integer programming formulations of the branch-and-bound variety have been reported by Geoffrion [5, 6], Kolesar [11], Petersen [15], and others. A "branch and search" algorithm with greatly reduced core storage and computer requirements has been developed by Greenberg and Hegerich [9]. The performance of general and special purpose integer programming computer codes suitable for the solution of (multiconstrained) knapsack problems have been reported in surveys by Trauth and Woolsey [19] and, more recently by Geoffrion and Marsten [7].

---

*Support received from the University of Wisconsin, Engineering Experiment Station and Graduate College is gratefully acknowledged.

Common to most approaches to the solution of knapsack problems is an exploitation of the one-to-one relationship between coefficients in the objective function and the constraint; leading to a particularly efficient ordering of the variables. In the multiconstraint problems considered in this paper such a mapping is not possible, and other approaches must be explored. Several modified branch-and-bound algorithms have successfully been applied to such problems [5, 6, 15, 20]. However, applications exist (Thesen [18]) where this approach leads to excessive computer time and/or computer storage requirements. This paper therefore presents an alternate approach to this problem resulting in greatly reduced demands for computer resources. Specifically, an enumerative algorithm with simple infeasibility tests, recursive bounding rules and systematic branching rules is being presented. Core storage requirements are reduced by applying a scheme similar to the one used by Greenberg and Hegerich [9], Zionts [20], and many others, resulting in a need to store only the current and one previous solution. Although a large number of alternatives are considered, improved execution times are achieved through the simplicity of the recursive algorithm.

The absence of any special, systematic structure in the constraints of the class of problems discussed in this paper makes them significantly harder to solve than many other classes of zero-one integer programming problems of comparable size. For example route selection and set covering problems both contain sets of mutually exclusive variables, thus greatly reducing the number of candidate variables to be considered for inclusion in a partially completed solution. It is frequently through implicit or explicit exploitation of this structure that authors (Geoffrion and Marsten [7], Salkin and Koncal [17], and others) are able to report on execution times superior to those reported in this paper.

II. THE PROBLEM

There are N projects. Each project has a fixed utility v_i. The combined utility of a set of projects is equal to the sum of the utilities of the individual projects. No penalty is invoked if a project is not chosen. Each project i requires a fixed amount, w_{ij} of K different resources. Only a limited amount, W_j, of each resource is available. If no more than L ($L \leq N$) of these projects may be chosen, find the combination of projects with the highest combined utility. (The upper bound on the number of projects is not always explicitly stated.) The integer programming formulation of this problem reads as follows:

$$\text{(1)} \qquad \text{Maximize } Z = \sum_{i=1}^{N} v_i x_i$$

$$\text{(2)} \qquad \text{Subject to } \sum_{j=1}^{N} w_{ij} x_j \leq W_i \qquad i = 1, \ldots, K$$

$$\text{(3)} \qquad \sum_{i=1}^{N} x_i \leq L$$

$$\text{(4)} \qquad x_i = 0 \text{ or } 1 \qquad i = 1, \ldots, N.$$

It is furthermore assumed that the variables are ordered such that $v_i \geq v_{i+1}$ and that all $w_i \geq 0$.

I. BRANCHING AND BOUNDING RULES

Murphree's Tree

The key to the current approach is the rigid structure of the particular tree being searched for ossible solutions. In order to discuss this structure and its ramifications, it is convenient to present otational concepts discussed by Knuth [10] and Murphree and Fenves [14] for efficient storage and nalysis of tree structures in a computer. Since a tree has no loops, it can always be drawn such that odes appear on discrete levels, and such that, for a directed tree, all nodes on the ith level point nodes on the $(i+1)$th level (Figure 1(a)). Such a tree can be redrawn as a binary tree when pointers ranches) of two different classes are introduced (Figure 1(b)).

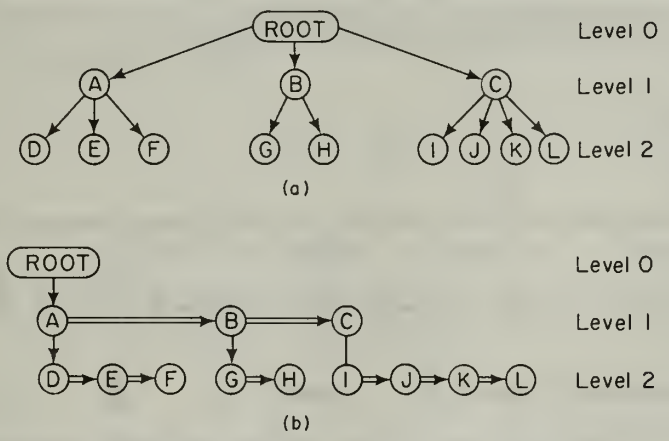


FIGURE 1. Directed tree (a) standard notation, (b) Murphree's notation

One pointer, the "First On Next" (FON) pointer, points to the *first* node in the set of successor odes on the *next level*. The other pointer, the "Next On Same" (NOS) Pointer, points to the *next* ode that belongs to the same set (the brother) as itself. Efficient algorithms for searching trees ganized in this manner are readily available.

The functions of the FON and the NOS pointers should be noted; the FON (successor) pointer eing somewhat similar to a conventional directed branch, while the NOS (brother) pointer may be terpreted as a graphical representation of a logical "or". Murphree introduces the double horizontal row (normally pointing to the right) as the NOS pointer, while a conventional downwards pointing row represents the FON pointer. This notation is adopted in Figure 1(b). Note that the two trees in gure 1 are isomorphic.

The Combinatorial Tree

A tree structure called the combinatorial tree can be used to define all combinations of objects at may be formed from a set of objects. The importance of this tree stems from the ease at which may be generated.

DEFINITION: A combinatorial tree of size L, N is a Murphree tree with L levels where the pointers om a node labelled "i" on the lth level point to nodes with the following labels.

Condition	*FON*	*NOS*
$l<L, i<N$	$i+1$	$i+1$
$l=L, i<N$	empty	$i+1$
$i=N$	empty	empty

where i, l, N, and L are positive integers, $L \leq N$, $1 \leq l \leq L$ and $1 \leq i \leq N$.

An example of a combinatorial tree (CT) with $L=3$ and $N=5$ is given in Figure 2. The importance of the combinatorial tree can be seen from the observation that there is a one to one correspondence between nodes and combinations. An obvious implication of this observation is the fact that a CT has $\sum_{1}^{L} \binom{N}{i}$ nodes. Indeed, the CT in Figure 2 has 25 nodes since 25 unique nonempty combinations of three or fewer objects can be drawn from a set of five objects.

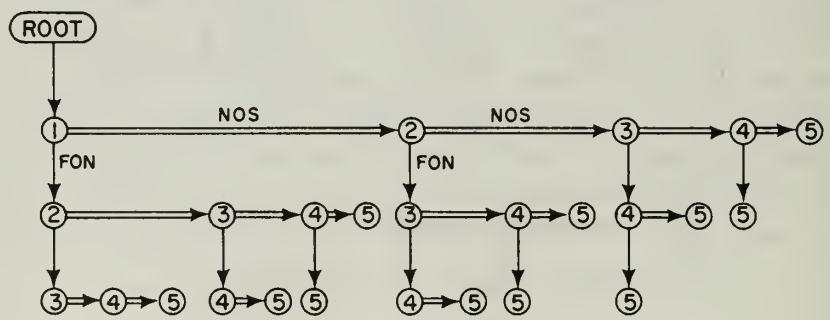


FIGURE 2. Combinatorial tree of size (3, 5)

It is clear that a CT of size L, N represents all feasible and many infeasible solutions to the problem defined in Equations 1 through 4, and that although the capacity constraints (2) may be violated, the zero-one constraints (4) never are.

The CT differs significantly from the binary enumeration tree applied by Balas [1, page 223]. While Balas' tree represents all variables regardless of their value, the CT includes only positive variables. The number of nodes in the tree is thus at least halved from $2^{N+1}-1$ to 2^N (usually more since L commonly is less than N). In the binary enumeration tree all combinations are represented by the *final* row of nodes. In fact Balas type algorithms complete partial solution by assuming variable values such that the final level is reached. On the other hand, since zeroes are not included in the CT all nodes represent unique solutions to a zero-one problem.

c. Upper Bounds on the Value of Potential Solutions

Each potential solution is represented by a (possibly incomplete) branch in the CT. By exploiting both the order in which variables appear along a branch and the upper limit on the number of nodes along a branch, an upper bound on the value of a solution may be obtained. Let

l = Number of variables in the current solution $(1 \leq l \leq L)$.

$R(i)$ = index of the ith variable in the current solution $(1 \leq i \leq l, R(i) < R(i+1))$.

$Z(l) = \sum_{i=1}^{l} v_{R(i)}$ (value of current solution).

$n = R(l)$ (last variable in the solution).

$B(l, n)$ = Upper bound on the value of the current solution.

This bound may be expressed as

$$B(l, n) = Z(l-1) + \sum_{i=n}^{\min(N, n+L-l)} v_i, \tag{5}$$

where min $(N, n+L-l)$ is the index of the last variable that might be included. It is easy to verify that (5) is an upper bound since $v_i > v_{i+1}$. This bound may be determined recursively from previous bounds through the following relationships:

$$B(1, 1) = \sum_{i=1}^{L} v_i \tag{6}$$

$$B(l, n+1) = B(l, n) - v_n + v_{n+L-l+1} \tag{7}$$

$$B(l+1, n+1) = B(l, n). \tag{8}$$

IV. THE ALGORITHM

The following enumerative algorithm utilizes the concepts discussed in the previous section to solve the problem defined in Equations 1 through 4. An overview of the enumerative nature of the algorithm is given in Figure 3.

STEP 1: *Initialization of Variables*

Define r_k as the current usage of the kth resource. Define Z^* as the value of the best solution yet found. Set $l, n = 0$; $v_i = 0$ for $i = N+1$ through $N+L$; $r_k = 0$ for $k = 1$ thru K. Set $B(1, 1) = \sum_{i}^{L} v_i$, Z and $Z^* = 0$. Compute a bound on Σx_i as min (g_j), where g_j is the maximum number of nonzero variables possible in the jth constraint. Set L equal to this bound unless a stricter bound is given. Proceed to Step 2.

STEP 2: *Add New Variable on New Level*

a. If $l = L$ or $n = N$, step is not possible, in this case proceed to step 4, otherwise continue to Step 2b.

b. Set $l = l+1$; $n = n+1$; $Z = Z + v_n$; $r_k = r_k + w_{n,k}$ for $k = 1$ through K; $R(l) = n$. Calculate bound on lateral step on this level from Equation (7): $B(l, n+1) = B(l, n) - v_n + v_{n+L-l+1}$. Proceed to step 3.

STEP 3: *Test Feasibility and Optimality of Solution*

a. If for at least one resource k, $r_k > W_k$: step was infeasible, in this case proceed to step 4, otherwise proceed to Step 3b.

b. If $Z > Z^*$ the current solution is best so far, in this case save the current solution by setting $Z^* = Z$, $l^* = l$, and $R^*(i) = R(i)$ for $i = 1$ through l^*. In either case attempt inclusion of other variables by proceeding to Step 2.

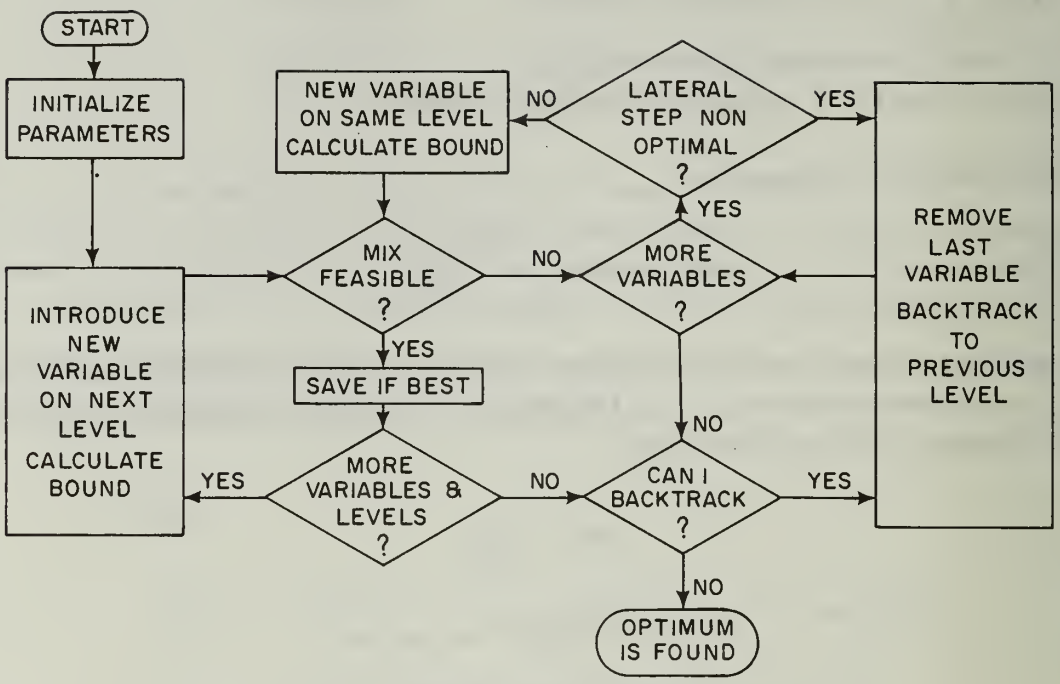


FIGURE 3. Overview of the algorithm

STEP 4: *Replace Variable on Current Level*

a. If $n=N$ step is not possible, in this case proceed to step 5, otherwise go to Step 4b.

b. If $B(l, n+1) < Z^*$ step will not yield an improved solution in this case proceed instead to Step 5.

c. Account for resource consumption by setting for all k

$$r_k = r_k - w_{n,k} + w_{n+1,k}$$

Also set $Z = Z - v_n + v_{n+1}$, $n = n+1$ and $R(l) = n$. Calculate bound on lateral step as

$$B(l, n+1) = B(l, n) - v_n + v_{n+L-l+1}.$$

Proceed to Step 3.

Step 5: *Backtrack to Previous Level*

a. If $l=1$ search is completed proceed to Step 6.

b. Remove variable on lth level by setting $Z = Z - v_n, r_k = r_k - w_{n,k}$ all k, $l = l-1$ and retrieve index of variable on this level as $n = R(l)$. Return to Step 4.

STEP 6: *Finish*

Optimal solution is contained in the first l^* elements of the vector R^*. Its value is Z^*

V. NUMERICAL EXAMPLE

As an illustration the following problem is solved:

$$\text{Maximize } Z = 10x_1 + 7x_2 + 4x_3 + 3x_4 + 1x_5$$

Subject to

$$2x_1 + 6x_2 + x_3 \leq 7$$

$$2x_1 + 3x_2 + 4x_3 + x_4 + x_5 \leq 8$$

$$x_1 + 2x_3 + 3x_4 + x_5 \leq 5$$

$$x_1 + x_2 + x_3 + x_4 + x_5 \leq 3$$

$$x_i = 0 \text{ or } 1.$$

Table 1 presents the detailed calculations for this example while the complete combinatorial tree is shown in Figure 2 and the tree traced is illustrated in Figure 4. Shaded nodes represent infeasible solutions, values along the double arrows represent bounds, and numbers adjacent to a node represent the value of the corresponding solution.

TABLE 1. *Steps in Calculations for Numerical Example*

Seq.	Comment	Stage	l	n	R	Z	r	Z^*	R^*	Bounds
1	Initialize	1	0	0	0	0	0, 0, 0	0	0	$B(1,1)=v_1+v_2+v_3=21$
2	Step down	2	1	1	1	10	2, 2, 1			$B(1,2)=B(1,1)-v_1+v_4=14$
3	Feasible and best	3						10	1	
4	Step down	2	2	2	1, 2	17	8, 5, 1			$B(2,3)=B(1,2)+v_1-v_2=17$
5	Not feasible	3								
6	Sidestep attractive	4		3	1, 3	14	3, 6, 3			$B(2,4)=B(2,3)-v_3+v_5=14$
7	Feasible and best	3						14	1, 3	
8	Step down	2	3	4	1, 3, 4	17	3, 7, 6			$B(3,5)=B(2,4)+v_3-v_5=15$
9	Not feasible	3								
10	Sidestep attractive	4		5	1, 3, 5	15	3, 7, 5			$B(3,6)=B(3,5)-v_5+v_6=15$
11	Feasible and best	3						15	1, 3, 5	
12	Cannot step down	2								
13	Cannot sidestep	4								
14	Backtrack	5	2	3	1, 3	14	3, 6, 3			
15	Sidestep unattractive	4								
16	Backtrack	5	1	1	1	10	2, 2, 1			
17	Sidestep unattractive	4								
18	Backtrack	5	0							
19	Optimum found	6								

VI. EVALUATION

a. Methodology

The algorithm was coded in FORTRAN and implemented and tested on a Datacraft 6024/3 mini computer with 32k 16-bit words. Since the algorithm is highly recursive and requires insignificant

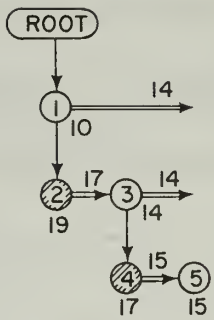


FIGURE 4. Tree searched for example problem

amounts of intermediate storage, the resulting code is short and requires minimum of core storage (10*K* words). Several hundred test problems have been solved using this code and two reference codes, the first based on the basic Balas algorithm (Geoffrion [5]) and the second on a recent (1972) extension of this algorithm (Zionts [19]). Attempts to implement Geoffrion's improved (1969) algorithm, RIP30C, (Geoffrion [6], Geoffrion and Nelson [8]) on the Datacraft were unsuccessful due to hardware restrictions. (The Datacraft requires two words for each floating point variable. Thus core requirements were excessive; also, sections of the RIP30C code requiring integer and floating point words to be of the same size could not be accommodated.)

A series of tests using randomly generated problems with controlled attributes was performed to determine both how various problem attributes affected the efficiency of the present algorithm, and how this code performed relative to the reference codes. Then, published test problems were used to provide reference data facilitating order of magnitude comparisons with evaluations performed elsewhere.

Most writers providing performance data for their and other algorithms point out the many uncertainties and difficulties in providing good comparison data. These problems include different levels of programmer skill, the absence of standard reference problems, differences in computer hardware, and the absence of generally accepted machine-independent comparison criteria. When placed in the proper context by specifying a computer and specific problem sets, however, the following information is useful when algorithms are compared and/or chosen for use in a particular application.

1. Solution time
2. Number of iterations required
3. Number of solutions saved
4. Amount of overshoot
5. Core requirements.

While items 1 and 5 are direct measures of the required computer resources, items 2, 3, and 4 provide machine-independent computer insight into the performance of the algorithm. In the results presented here the iteration counter was incremented whenever Stage 3b of the algorithm was reached.

b. Test Problems

A preliminary study was made to determine those attributes of the problem at hand that had the most significant impact on solution time.

The following parameters were identified:

1. Problem size (K and M)
2. Constraint density ($D = Pr\{w_{ij} > 0\}$)
3. Constraint tightness ($T = \Sigma(\Sigma w_{ij}/W_i)$)
4. Difference between adjacent v_j's.

Capital budgeting problems of the Lori-Savage type typically have a density close to unity and a tightness of between one and two, while set covering problems on the other hand may have densities as low as 0.02–0.11 and a corresponding tightness of 2 to 15 depending on problem size.

Results

Randomly generated problems with controlled attributes were solved to determine the effect of these attributes on solution time. To keep cost down, the size of these problems was, whenever possible, kept to 5×15. Spot checks indicate that the trends developed for these problems do extrapolate to problems of larger size.

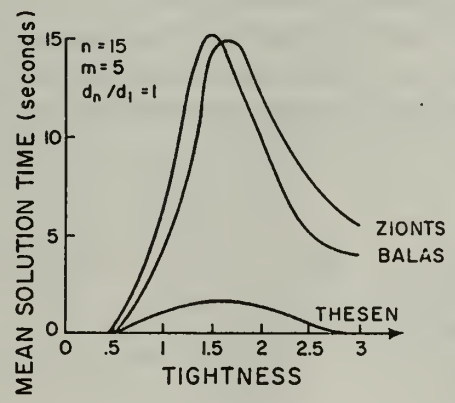


FIGURE 5. Effect of constraint tightness on solution time

Constraint Tightness.—Performance of the present algorithm and the two reference algorithms showed a strong dependence upon constraint tightness. This is shown in Figure 5 where the relationship between solution time and constraint tightness for the case with evenly spaced v_j's. It is seen that solution time for all three algorithms increased as the tightness was increased from 0.5 to approximately 1.5–1.6. Further increases caused a corresponding reduction in solution time. For relatively unconstrained problems the Balas-Geoffrion code was 20- to 30-percent faster than the Zionts code. On the other hand, for more constrained problems ($t > 1.6$) the Zionts code was 20- to 30-percent more efficient than the other code. This behavior is explained by the fact that Zionts' extension of the Balas algorithm involves an exploitation of bounds on variables induced by tight constraints. For relatively unconstrained problems the additional work required by this approach is not offset by the benefits gained therefrom. The present code showed the same sensitivity to constraint tightness with problems in the tightness range of 1 to 2.5 again being the hardest to solve. However, as a rule of thumb, this code required approximately one-tenth of the solution time required by the two reference codes for problems of this size.

Objective Function.—The three codes also exhibited a strong sensitivity to the relative values of

the terms in the objective functions. For example, the average solution time for 10 randomly generated five-constraint 15-variable problems with $t=2$ and equal v_j's was 18.36, 24.03, and 2.13 seconds, respectively, for the Balas, Zionts, and Thesen codes. These times dropped exponentially as the ratio between the largest and the smallest of coefficient in the objective function (d_n/d_1) was increased. For $d_n/d_1=5$ the times were 6.63, 5.53, and 0.12 sec. Similar trends were found for other problem sizes and tightnesses. The relative effect of the objective function on solution time appeared to be independent of problem tightness and similar for all three codes.

Problem Size.—Random problems of successively larger size were generated and solved to determine the effect of problem size on solution time. Figure 6 shows mean solution times for the present code and the two reference codes for random problems with a tightness of 1.0 and $d_n/d_1=1$. It is seen that the magnitude and rate of increase in solution times is substantially less for the present code than for the two reference codes. However, it is still increasing as the problem time increases. The rate of increase in solution time was found to be dependent upon constraint tightness. While on the average 1 min was required to solve a 30-variable problem with $t=2$, a 50-variable problem could be solved in the same time if the tightness was reduced to 1.

The code was also used to solve randomly generated inequality set covering problems. As in [17] a density of 0.07 and a d_n/d_1 of 100 was used. The results are given in Table 2. While solution times for this class of problems are less than those of the knapsack type, the increasing rate of increase

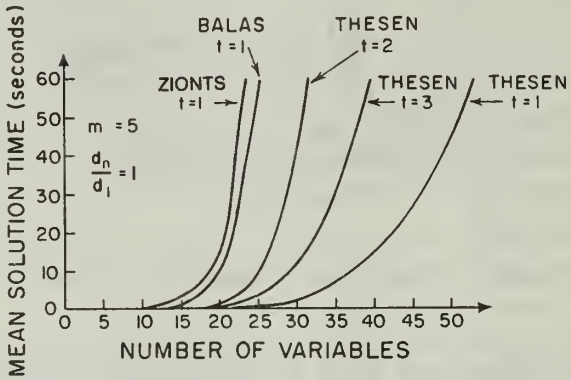


FIGURE 6. Problem size

TABLE 2. *Computational Results for Randomly Generated Setcovering Problems**

K	N	L	Iterations	Solutions saved	Time to optimal found	Total time	Overshoot
30	10	10	10.5	7.25	0.03	0.05	16.6
30	20	15	106.5	12.75	0.052	0.54	10.4
30	30	20	1,686.5	19	2.23	9.93	4.4
30	40	30	14,039.3	20	9.62	87.60	9.1
15	10	10	10.75	6.75	0.01	0.03	3.0
15	20	15	123.25	15	0.05	0.30	6.0
15	30	20	927.25	20.25	0.32	2.74	8.6
15	40	30	22,792.5	25.75	2.25	62.64	27.9

* Averages from four independently generated problems.

in solution time with problem size is again observed. Table 2 also illustrates that while the present algorithm finds the optimum solution reasonably fast, an excessive amount of time is consumed before this optimum is verified.

The algorithms were also used to solve Petersen's seven capital budget problems [14] This result is shown in Table 3.

Bounds. — Tests were made to determine the effect of the bound on the number of nonzero variables on solution speed. Figure 7 shows this effect on Petersen's problem number 5. It is seen that a tight bound (or accurate estimate) on the number of nonzero variables in the optimal solution will play an important role in reducing the computer time required to solve a problem. This property will be particularly useful when large problems with relatively few nonzero variables are to be solved.

Most branch and bound type algorithms exhibit improved performance when a tight bound on the value of the optimal solution is known. Such a bound cannot be exploited directly in the present

TABLE 3. *Results From Seven Problems Introduced by Petersen* [15]

Problem No.	Constraints	Variables	Thesen					Balas/ Geoffrion [5]
			L	Total Time	Time To Optimum	Iterations ($L=N$)	Solutions Saved	
1	10	6	6	0.007 sec				0.2 sec
2	10	10	10	0.011 sec				0.5 sec
3	10	15	15	0.041 sec				1.9 sec
4	10	20	20	0.085 sec	0.02	272	9	20.9 sec
5	10	28	28	2.70 sec	1.28	1,538	20	309.8 sec
6	5	39	27	308.54 sec	298.43	263,589	91	>360 sec
7	5	50		>360 sec				>360 sec

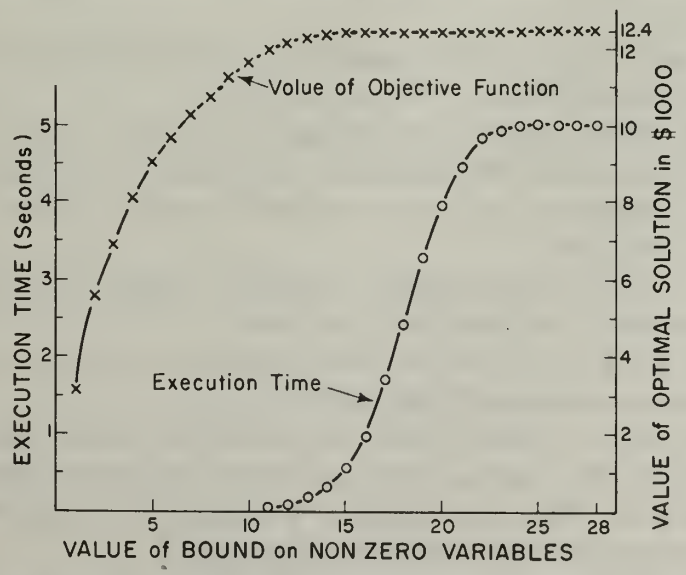


FIGURE 7. Effect of the bound on the number of nonzero variables on execution time for Petersen's problem 5

algorithm. This bound may, however, be used to calculate a bound on the number of nonzero variables, thus indirectly resulting in improved performance.

VII. CONCLUSION

A data structure forming a convenient basis for the development of enumerative solution algorithms for combinatorial problems has been presented. With this data structure as a basis, a recursive algorithm for the solution of the multidimensional knapsack problem was developed. This algorithm was shown to be able to solve reference problems in substantially shorter time than two other algorithms. This improved speed is primarily based in the algorithms recursive simplicity, allowing for an examination of a significantly larger number of alternative solutions per unit of time.

While the algorithm presented here is competitive for small to medium sized problems, or large problems with few nonzero variables, applying this algorithm to the solution of large (100 variables) project selection problems may be impractical. It should, however, be pointed out that should it be necessary to use another code to solve these large problems, it will be at the cost of substantially increased core and off line storage requirements.

From an analysis of the performance data, it is evident that the algorithm finds optimal solutions reasonably fast, even for larger problems, but that an unusually large amount of time is then spent verifying this solution. Improved performance can therefore, be expected when tighter bounds on the value of an alternative solution (eg. 8) is developed. Preliminary studies of several special structure problems (such as set covering problems) indicate that the current algorithm allows the constraint structure of the problem to be explicitly exploited to develop such improved bounds.

BIBLIOGRAPHY

[1] Balas, E., "An Additive Algorithm for Solving Linear Programs with Zero-One Variables," Operations Research *13*, 517–49 (July–Aug. 1965).

[2] Bellman, R. E. and S. E. Dreyfus, *Applied Dynamic Programming* (Princeton University Press, 1962), pp. 27–31.

[3] Feuerman, M. and H. Weiss, "A Mathematical Programming Model for Test Construction and Scoring," Management Science *19*, 961–66 (Apr. 1973).

[4] Fulkerson, D. R., "Flows, Networks and Combinatorial Research," RM 3378, The Rand Corporation, Santa Monica, Calif., 1962, reprinted in Geoffrion, A. M., *Perspectives of Optimization.*

[5] Geoffrion, A. M., "Integer Programming by Implicit Enumeration and Balas' Method," SIAM Review *9*, 178–190 (Apr. 1967).

[6] Geoffrion, A. M., "An Improved Implicit Enumeration Approach for Integer Programming," Operations Research *17*, 437, 454 (May–June 1969).

[7] Geoffrion, A. M. and R. E. Marsten, "Integer Programming: A Framework and State-of-the-Art Survey," Management Science *18*, 465–491 (1972).

[8] Geoffrion, A. M. and A. B. Nelson, "Users Instructions for Integer Linear Programming Code RIP30C; RM–5627–PR., The Rand Corporation, Santa Monica, Calif. (1968).

[9] Greenberg, Harold and Robert L. Hegerich, "A Branch Search Algorithm for the Knapsack Problem," Management Science *16*, 327–332 (Jan. 1970).

[10] Knuth, D., *The Art of Computer Programming* (Addison-Wesley, 1968), vol. 1.

[11] Kolesar, Peter J., "A Branch and Bound Algorithm for the Knapsack Problem," Management Science *9*, 497–520 (May 1967).

[12] Kolesar, Peter J., "Linear Programming and the Reliability of Multicomponent Systems," Nav. Res. Log. Quart. *14*, 317–328 (Sept. 1967).

[13] Land, A. H. and A. G. Doig, "An Automatic Method for Solving Discrete Programming Problem," Econometrica *28*, 497–520 (1960).

[14] Murphee, E. L. and S. Fenves, "A Technique for Generating Interpretive Translators for Problem Oriented Languages" BIT (Oct. 1970).

[15] Petersen, Clifford C., "Computational Experience with Variants of the Balas Algorithm Applied to the Selection of R&D Projects," Management Science *13*, 736–750 (May 1967).

[16] Pierce, J. F. and J. S. Lasky, "Improved Combinatorial Algorithms for All Zero-one Integer Programming," Management Science *19*, 528–543 (Jan. 1973).

[17] Salkin, Harvey M. and Ronald D. Koncal, "Set Covering by an All Integer Algorithm: Computational Experience," Journal of the ACM *20*, 189–193 (Apr. 1973).

[18] Thesen, Arne, "Scheduling of Computer Programs in a Multiprogramming Environment," BIT *13–2*, 208–216 (1973).

[19] Trauth, C. A. and R. E. Woolsey, "Integer Linear Programming: A Study in Computational Efficiency," Management Science *15*, 481–493 (May 1969).

[20] Zionts, Stanley, "Generalized Implicit Enumeration Bounds on Variables for Solving Linear Program with Zero-One Variables," Nav. Res. Log. Quart. *19*, 165–181 (1972).

PARAMETRIC STUDIES IN TRANSPORTATION-TYPE PROBLEMS

C. O. Fong
University of Malaya
Kuala Lumpur, Malaysia

and

M. R. Rao
University of Rochester

ABSTRACT

In this paper we present some results in parametric studies on several transportation-type problems. Specifically, a characterization is obtained for the optimal values of the variables in the problem of determining an optimal growth path in a logistics system. We also derive an upper bound beyond which the optimal growth path remains the same. The results are then extended to the goal programming model and the prespecified market growth rate problem.

1. INTRODUCTION

Post optimization procedures for the transportation problem have been studied by Balas and Ivanescu [1], Charnes and Cooper [2, 3], and Grabowski and Szwarc [9]. Srinivasan and Thompson [14] give an algorithm for solving the problem of finding an optiaml growth path in a logistics system. Several extensions including the goal programming model and the prespecified market growth rate problem are also treated in [14].

In this paper the problem of determining an optimal growth path in a logistics system is studied more directly as a transportation-type network problem. In section 2 we give an interesting characterization of the values of the variables in an optimal growth path. Further we derive an upper bound beyond which the optimal growth path does not change. The goal programming model is also discussed in the same section. The results are extended to the prespecified market growth rate problem in section 3. The last section contains some concluding remarks.

2. LOGISTICS OPERATIONS PROBLEM

The optimal growth path problem $(P1)$ in a logistics system has been formulated in [14] as:

1) $$Z(\delta) = \text{Min} \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{i \in I} g_i S_i + \sum_{j \in J} h_j T_j,$$

2) $$\sum_{j \in J} x_{ij} - S_i = q_i \qquad \text{for } i \in I = \{1, 2, \ldots m\}, \text{ the set of sources,}$$

3) $$\sum_{i \in I} x_{ij} - T_j = r_j \qquad \text{for } j \in J = \{1, 2, \ldots n\}, \text{ the set of markets,}$$

$$\sum_{i\in I}\sum_{j\in J} x_{ij}=K_0+\delta, \text{ and} \tag{4}$$

$$\sum_{i\in I} q_i=\sum_{j\in J} r_j=K_0, \tag{5}$$

where

$q_i(r_j)$ = minimum amount to be supplied from source i (to market j),
$S_i(T_j)$ = surplus amount supplied from source i (to market j),
$0 \leq g_i(h_j)$ = unit cost of supplying from source i (to market j),
x_{ij} = amount of goods shipped from source i to market j, and
c_{ij} = unit cost of shipment from source i to market j.

The optimal growth path problem [1] consists of studying the variations in the optimal values of Z, $\{x_{ij}\}$, $\{S_i\}$, and $\{T_j\}$ as a function of δ. Utilizing the methodology developed in [13], a procedure for tracing the optimal growth path is given in [14]. In this section, we convert $P1$ into its equivalent network representation and from the structure of this network derive some interesting characterizations of the optimal values of $\{X_{ij}\}$, $\{S_i\}$, and $\{T_j\}$.

Summing (3) over J, and then subtracting this sum from (4), we have in place of (4) an equivalent constraint:

$$\sum_{j\in J} T_j=\delta\,. \tag{6}$$

Let (V, E) be the network of a $m \times n$ transportation problem [6, 12]. By adding the redundant constraint $\sum_{i\in I} S_i=\delta$, a network representation of $P1$ can now be seen to be an extension of (V, E) with the addition of a supersource s and a supersink t together with arc (s, i) for $i\in I$ and arc (j, t) for $j\in J$. Each S_i is represented by the flow through (s, i). Similarly, each T_j is represented by the flow through (j, t). There is a supply of δ at node s and a requirement of δ at node t. Let this extended network be (V_1, E_1). It should be noted that in (V_1, E_1) there is a fixed supply (demand) of $a_i(b_j)$ at each node $i(j)$. The optimal growth path problem now consists of tracing the optimal flow through (V_1, E_1) as a function of total flow out of t.

Consider the following parametric procedure to solve $P1$. For $\delta=0$ find a basic optimal solution to $P1$. Corresponding to this basic optimal solution we have a unique path Γ from s to t. Now, try to increase δ until one of the variables on the path Γ becomes zero. Choose a variable [2] on the path Γ with value zero as the variable to leave the basis. Now perform a dual simplex iteration and try to increase δ further until some other variable on the path Γ becomes zero. Continue this process until δ

[1] The problem becomes trivial if $K_0=0$ or if $|I|$ or $|J|=1$. We assume here that $K_0>0$ and that $|I|$ and $|J| \geq 2$.

[2] If there are several variables on the path Γ with value zero, the following rule avoids the possibility, if any, of cycling. The tree defined by the basis is rooted at node s. Define the distance of a basic variable corresponding to arc (k, l) as the number of arcs in the tree between nodes s and l. Among the variables on path Γ with value zero, choose the one which is closest (minimum distance) to node s. See also Minieka [10].

basis is obtained which is optimal for all $\delta \geq \bar{\delta}$ where $\bar{\delta}$ is a constant. Existence of such a basis is guaranteed by a well known result in parametric linear programming; see for example [11, chap. VIII].

Using this procedure, we can easily prove the result that there exists an optimal solution to $P1$ such that the optimal values of $\{S_i\}$ and $\{T_j\}$ are nondecreasing functions of $\delta \geq 0$ (Theorem 2 of [14]). In addition, we have the following result.

THEOREM 1: Suppose for $\delta = \delta^*$, there exists a basic optimal solution to $P1$ with S_u (T_v) basic. Let x^*_{uj} for $j \in J$ (x^*_{iv} for $i \in I$) represent the optimal flow values corresponding to $\delta = \delta^*$. Then there exists an optimal solution to $P1$ such that x_{uj} for $j \in J$ (x_{iv} for $i \in I$) is a nondecreasing function of $\delta \geq \delta^*$ with $x_{uj} = x^*_{uj}$ ($x_{iv} = x^*_{iv}$) for $\delta = \delta^*$.

PROOF: Corresponding to $\delta = \delta^*$ let B be an optimal basis and S_u be a basic variable. For $\delta \geq \delta^*$ S_u is nondecreasing and hence remains in the basic set. Now we show that $x_{uj}, j \geq J$ is nondecreasing as δ is increased from δ^*. Let Γ be the unique path from s to t as defined by B. When δ is increased from δ^*, the flow along Γ is changed correspondingly. If arc (s, u) is in Γ, since B defines a tree, there must be precisely one arc (u, r) $r \in J$ in Γ. Thus x_{ur}, $r \in J$ increases while $x_{uj}, j \in J - \{r\}$ remains unchanged. On the other hand if arc (s, u) is not in Γ, no arc (u, r), $r \in J$ can be in Γ for otherwise we would have a cycle with arc (s, u). In this case all x_{uj}, $j \in J$ remain unchanged. Thus as δ is increased from δ^*, x_{uj}, $j \in J$ is nondecreasing. Repeating the same argument for any $\delta \geq \delta^*$, we have x_{uj}, $j \in J$ nondecreasing for $\delta \geq \delta^*$ with $x_{uj} = x^*_{uj}$ for $\delta = \delta^*$. A similar argument provides the same result for x_{iv}, $i \in I$ if T_v is a basic variable.

This result is interesting and can have managerial implications. It says that after a source is required to provide a surplus amount into the system, each optimal shipment out of that source is nondecreasing as the system continues to grow from that point in time. Similarly, after a market is required to accept a surplus amount, each optimal shipment into that market is nondecreasing as the system continues to grow from that point in time.

2.1. Bound for δ

$P1$ is feasible for any value of $\delta \geq 0$, and by a well known result in parametric linear programming, there exists a value $\bar{\delta}$ and an optimal basis $\bar{B}$ which remains optimal for all $\delta \geq \bar{\delta}$. In this section we derive an upper bound $\hat{\delta}$ for $\bar{\delta}$, such that independent of the costs, an optimal basis $\hat{B}$ remains optimal for all $\delta \geq \hat{\delta}$. For ease of presentation we derive the bound under the assumption all q_i and $r_j > 0$. If this is not true we could consider the perturbed problem in which all q_i and $r_j > 0$.

Let B be any feasible basis to $P1$ (with the constraint set (2), (3), and (6)), B^{-1} be its inverse and a_j be any column of the constraint set. Let β_{ij} be the element in the ith row and jth column of B^{-1}. Every basis B has $(m+n+1)$ rows and columns. Let B be permuted such that its first m rows are derived from (2), its $(m+1)$ to $(m+n)$ rows are derived from (3) and the last row from (6). B^{-1} can correspondingly be partitioned into three sets of columns, the first set ($N_1 = \{1 \ldots m\}$) consisting of the first m columns, the second set ($N_2 = \{m+1, \ldots m+n\}$) consisting of the $(m+1)$ to $(m+n)$ columns and the third set ($N_3 = \{m+n+1\}$) the $(m+n+1)$ column. Before deriving $\hat{\delta}$, we present below several propositions regarding the structure of B^{-1}. Their proofs are relatively straightforward and are given in [7].

PROPOSITION 1: For any row i of B^{-1}, either (i) $\beta_{ij} = 0$ or 1 for all $j \in N_1$ and $\beta_{ij} = 0$ or -1 for all $j \in N_2$ or (ii) $\beta_{ij} = 0$ or -1 for all $j \in N_1$ and $\beta_{ij} = 0$ or 1 for all $j \in N_2$.

PROPOSITION 2: For any row i of B^{-1}

(a) If $\beta_{ij} = -1$ for any $j \epsilon N_1(N_2)$, then there exists at least one $v \epsilon N_2(N_1)$ with $\beta_{iv} = 1$.

(b) If $\beta_{ij} = -1$ for all $j \epsilon N_1$, then $\beta_{iu} = 1$ for all $u \epsilon N_2$.

PROPOSITION 3: For each row i of B^{-1}, (a) there exists a $i \epsilon N_1$ such that $\beta_{ij} = 0$ or -1, (b) there exists a $j \epsilon N_2$ such that $\beta_{ij} = 0$ or 1.

We are now in a position to derive the upper bound $\hat{\delta}$.

THEOREM 2: Let $\hat{B}$ be an optimal basis to $P1$ for all $\delta \geq \bar{\delta}$. Then $\hat{B}$ represents an optimal basis to $P1$ for all $\delta \geq \hat{\delta}$ where $\hat{\delta} = K_0 - q_u - r_v$ and $q_u = \underset{i \epsilon I}{\text{Min}} \{q_i\}$, $r_v = \underset{j \epsilon J}{\text{Min}} \{r_j\}$.

PROOF: Let $d(\delta) = [q_1, \ldots, q_m, r_1, \ldots, r_n, \delta]^T$ denote the right hand side vector of $P1$. From Proposition 1 the kth element of $\hat{B}^{-1} d(\delta)$ must be in the form of one of the following:

$$(7) \qquad \sum_{i \epsilon I_k^+} q_i - \sum_{j \epsilon J_k^-} r_j \geq 0 \qquad \text{or} - \sum_{i \epsilon I_k^-} q_i + \sum_{j \epsilon J_k^+} r_j \geq 0,$$

$$(8) \qquad \sum_{i \epsilon I_k^+} q_i - \sum_{j \epsilon J_k^-} r_j - \delta \geq 0 \qquad \text{or} - \sum_{i \epsilon I_k^-} q_i + \sum_{j \epsilon J_k^+} r_j - \delta \geq 0,$$

$$(9)^{(3)} \qquad \text{(a)} \sum_{i \epsilon I_k^+} q_i - \sum_{j \epsilon J_k^-} r_j + \delta \geq 0 \quad \text{or (b)} - \sum_{i \epsilon I_k^-} q_i + \sum_{j \epsilon J_k^+} r_j + \delta \geq 0,$$

where

$$I_k^+ = \{j | \beta_{kj} = 1, j \epsilon N_1\}, I_k^- = \{j | \beta_{kj} = -1, j \epsilon N_1\}$$
$$J_k^+ = \{j | \beta_{k, j+m} = 1, j + m \epsilon N_2\}, J_k^- = \{j | \beta_{k, j+m} = -1, j + m \epsilon N_2\}$$

and satisfy the conditions

(a) If $I_k^- (J_k^-) \neq \phi$, then $J_k^+ (I_k^+) \neq \phi$ (Proposition 2(a)),

(b) If $I_k^- = I$ then $J_k^+ = J$ (Proposition 2(b)), and

(c) $I_k^+ \neq I$ $(J_k^- \neq J)$ (Proposition 3(a), (b)).

Since B is an optimal basis for all $\delta \geq \bar{\delta}$, the kth term of $B^{-1} d(\delta)$ must be given by (7) or (9)(a) or (b). Clearly if the kth term of $\hat{B}^{-1} d(\delta)$ is given by (7) its value is ≥ 0 for all $\delta \geq 0$ and hence in particular for $\delta \geq \hat{\delta}$. Suppose the kth term is given by (9) (b). Now, if $I_k^- = I$ then $J_k^+ = J$ and consequently the value of the kth term is given by δ which is ≥ 0 for all $\delta \geq 0$. Let $I_k^- = I - \{u\}$ and $J_k^+ = \{v\}$. It is easily seen that I_k^-, J_k^- satisfy conditions (a), (b), and (c) and furthermore provide an upper bound $\hat{\delta} = K_0 - q_u - r_v$ for δ. Since $J_k^- \neq J$ and $J_k^- \neq \phi$ implies that $I_k^+ \neq \phi$, the same upper bound is obtained from (9)(a) also.

---

(3) It can be shown that in (9)(a) $J_k^- = \phi$, but this is not essential for the proof of Theorem 3.

Consequently, for all $\delta \geq \hat{\delta} = K_0 - q_u - r_v$, $B^{-1}d(\delta)$ is feasible and hence optimal to $P1$. This completes the proof of Theorem 2.

It is evident that for $\delta \geq \hat{\delta}$, in the network (V_1, E_1) the unique path Γ from s to t as defined by $\hat{B}$ is of the form:

(10) $$\Gamma = \{(s, i), (i, j), (j, t)\}.$$

Theorem 2 and Equation (10) provide an important characterization of the optimal growth path of $P1$. From Theorem 2, we know that for $\delta \geq K_0 - q_u - r_v$, the optimal growth path becomes invariant (i.e., the same basis remains optimal). Further from (10), this optimal growth path is Γ, which means that all increases in the system shipment $\left(\sum_{i \in I} \sum_{j \in J} x_{ij} \right)$ beyond $\hat{\delta}$ is shipped from a single source i to a single market j, through route (i, j). The surprising aspect of Theorem 2 is that the upper bound on $\hat{\delta}$ is so low.

Although in general $\bar{\delta} \leq \hat{\delta}$, it is easy to construct an example in which $\bar{\delta} = \hat{\delta}$ (see [7]). It is well known [3], [11] that $Z(\delta)$ is a piecewise linear convex function of δ. Clearly, the slope of this function for $\delta \geq \hat{\delta}$ is given by $\underset{(ij)}{\text{Min}}\, g_i + c_{ij} + h_j$. It can be easily shown by an example that $Z(\delta)$ may exhibit the "more for less" paradox (see also [4], [5], and [15]).

2.2. Goal Programming Model

A problem ($P2$) equivalent to the goal programming model in [14] is $P1$ with (1), (2), (3), and (6) replaced respectively by

(1a) $$Z(\delta) = \text{Min} \sum_{i \in I} \sum_{j \in J} c_{ij}x_{ij} + \sum_{i \in I} g_i S_i + \sum_{j \in J} h_j T_j + \sum_{i \in I} \phi_i w_i + \sum_{j \in J} \sigma_j y_j .$$

(2a) $$\sum_{j \in J} x_{ij} - S_i + w_i = q_i \qquad i \in I$$

(3a) $$\sum_{i \in I} x_{ij} - T_j + y_j = r_j \qquad j \in J, \text{ and}$$

(6a) $$\sum_{j \in J} T_j - \sum_{j \in J} y_j = \delta,$$

where $w_i(y_j)$ represents the slack at the ith supply (jth market), and ϕ_i (σ_j) represents the penalty cost per unit of slack at the ith supply (jth market). $P2$ is bounded if and only if we have $-g_i \leq \phi_i$ for all $i \in I$ and $-h_j \leq \sigma_j$ for all $j \in J$. Henceforth we assume this to be the case.

A network representation (V_2, E_2) of $P2$ is given by (V_1, E_1) with the additional arcs (i, s), $i \in I$ and (t, j), $j \in J$. Proposition 3 is now valid only for feasible basis that correspond to $\delta > 0$ while the other propositions remain valid although the proofs in some cases are different. Theorems 1 and 2 are valid for $P2$ also. In addition, it can be established that there exists and optimal solution to $P2$ such that the optimal values of $\{S_i\}$ and $\{T_j\}$ are nondecreasing functions of $\delta \geq 0$ while $\{w_i\}$ and $\{y_j\}$ are non-

increasing functions of $\delta \geq 0$ (see [7] for details). Furthermore, for $\delta = \delta^*$, if $S_u(T_v)$ is in the optimal basis, then there exists an optimal solution to $P2$ with $w_u = 0$ ($\gamma_v = 0$) for all $\delta \geq \delta^*$. If all ϕ_i, $i \in I$ and $\sigma_j, j \in J$ are positive and sufficiently large, clearly $P2$ becomes equivalent to $P1$, and we could still have the "more for less" paradox. However, if $\phi_i \leq 0$ for all $i \in I$ and $\sigma_j \leq 0$ for all $j \in J$ it can be shown (see [7]) that in $P2$, $Z(\delta)$ has only nonnegative slopes.

3. PRESPECIFIED MARKET GROWTH RATES

The prespecified market growth rates problem ($P3$) can be formulated as:

$$Z(\delta) = \text{Min} \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} + \sum_{i \in I} g_i S_i, \tag{11}$$

$$\sum_{j \in J} x_{ij} - S_i = q_i \qquad \text{for } i \in I, \tag{12}$$

$$\sum_{i \in I} x_{ij} = r_j + f_j(\delta) \qquad \text{for } j \in J, \text{ and} \tag{13}$$

$$x_{ij}, S_i \geq 0 \qquad \text{for } i \in I, j \in J, \tag{14}$$

where $\sum_{i \in I} q_i = \sum_{j \in J} r_j = K_0$ and $f_j(\delta), j \in J$ are the given market growth rates with each $f_j(\delta)$ a continuous piecewise linear nondecreasing function of δ and $f_j(0) = 0, j \in J$. $P3$ is more general than the formulation in [14] because it permits arbitrary growth rates rather than just arithmetic growth. The problem is to find the optimal values of Z, $X = \{x_{ij}\}$ and $\{S_i\}$ as a function of δ. By adding the redundant constraint $\sum_{i \in I} S_i = \sum_{j \in J} f_j(\delta)$ a network representation of $P3$ is given by (V_1, E_1) with arcs (j, t) for $j \in J$ removed, and each demand node $j \in J$ now has a demand of $r_j + f_j(\delta)$ instead of r_j. Let this network be (V_3, E_3).

We next provide some interesting characterizations of the optimal $X = \{x_{ij}\}$ and $\{S_i\}$ as a function of δ. Before we do that we need some further notation.

Let $d(\delta) = (q_1, \ldots q_m, r_1 + f_1(\delta), \ldots r_m + f_m(\delta))^T$ be the right hand side vector of $P3$.

Let $R_j = \{\delta | f_j(\delta)$ is a break point$\}$ $j \in J$, $R = \bigcup_j R_j = \{\delta_1, \delta_2, \ldots \delta_p\}$ where $\delta_i < \delta_{i+1}$ for $i = 1, \ldots, p-1$.

Let $\zeta_{jk} \geq 0$, $k = 1, \ldots, p+1$; $j \in J$ be the slope of $f_j(\delta)$ in the interval $\delta_{k-1} < \delta < \delta_k$ where $\delta_0 = 0$ and $\delta_{p+1} = \infty$.

$$d_1(\delta) = (q_1, \ldots, q_m, r_1 + \zeta_{11}\delta, \ldots, r_m + \zeta_{m1}\delta) \qquad \text{for } 0 \leq \delta \leq \delta_1.$$

$$d_k(\delta)=\Big(q_1, \ldots, q_m, r_1+\sum_{l=1}^{k-1}\zeta_{1l}(\delta_l-\delta_{l-1})+\zeta_{1k}(\delta-\delta_{k-1}), \ldots, r_m$$

$$+\sum_{l=1}^{k-1}\zeta_{ml}(\delta_l-\delta_{l-1})+\zeta_{mk}(\delta-\delta_{k-1})\Big) \qquad \text{for } k=2, \ldots, p+1$$

$$\text{and } \delta_{k-1}\leq\delta\leq\delta_k.$$

Note that $d_k(\delta_k)=d_{k+1}(\delta_k) \qquad k=1, 2, \ldots p.$

THEOREM 3: There exists an optimal solution to $P3$ such that the optimal values of $\{S_i\}$ are nondecreasing functions of $\delta \geq 0$.

PROOF: Initially let $d_1(\delta)$ be the right hand side of $P3$. Consider the following parametric procedure to solve $P3$. First find a basic optimal solution to $P3$ with $\delta = 0$. Corresponding to this basic optimal solution we have a unique path Γ_j from s to $j\epsilon J$. Now try to increase δ until either $\delta = \delta_1$ or one of the variables on a path Γ_j, $j\epsilon J$ becomes zero. If a variable on the path Γ_v becomes zero choose that particular variable [4] to leave the basis. Now perform a dual simplex iteration to obtain a new basis and try to increase δ until some other variable on a path Γ_j, $j\epsilon J$ becomes zero. If $\delta = \delta_1$, let $d_2(\delta)$ be the right hand side of $P3$ instead of $d_1(\delta)$. More generally, if $\delta = \delta_k$ (for $k \leq p$) let $d_{k+1}(\delta)$ be the right hand side instead of $d_k(\delta)$. The process is repeated until $d_{p+1}(\delta)$ is the right hand side of $P3$ and we obtain a basis which is optimal for all $\delta \geq \overline{\delta} \geq \delta_p$. Existence of such a basis is guaranteed by known results in parametric linear programming. For any given value of δ, let Γ_j be the unique flow augmenting path from s to $j\epsilon J$ in the network (V_3, E_3). Clearly Γ_j includes exactly one arc of the form (s, i_j), $i_j\epsilon I$. As δ is increased, the optimal values of S_{ij}, $i_j\epsilon I$, $j\epsilon J$ are increasing while the remaining S_k are unchanged. Consequently, after δ is increased the basic variables S_u do not leave the basis. The entire parametric procedure is such that once a variable S_u enters the basis, it never leaves the basis. Furthermore, the values of S_u are nondecreasing. Thus the theorem follows.

THEOREM 4: Suppose for $\delta=\delta^*$, there exists a basic optimal solution to $P3$ with S_u in the basis. Let x^*_{uj} for $j\epsilon J$ represent the optimal values corresponding to $\delta=\delta^*$. Then there exists an optimal solution to $P3$, such that x_{uj} for $j\epsilon J$ are nondecreasing functions of $\delta \geq \delta^*$ with $x_{uj}=x^*_{uj}$ for $\delta=\delta^*$.

This theorem can be proved by arguments similar to those given for Theorems 1 and 3. The details are omitted here.

The results of Theorems 3 and 4 are interesting. From Theorem 3 we know that the optimal surplus addition in each source i is a nondecreasing function of δ. Further, from Theorem 4 we know that beyond the point in time at which there is a surplus addition at a source i, each individual optimal shipment out of that source is a nondecreasing function of δ.

3.1. Bound for δ

We now derive an upper bound $\hat{\delta}$ for δ such that the same basis is optimal to $P3$ for $\delta \geq \hat{\delta}$. For ease of presentation we provide the bound under the assumption [5] that all q_i and $r_j > 0$.

---

[4] As mentioned in footnote 2, cycling due to degeneracy can be avoided by choosing the variable which is closest to node s and is on some path Γ_j, $j\epsilon J$ with value zero. Ties may be broken arbitrarily.

[5] As pointed out in section 2.1, if this is not true we could consider the perturbed problem in which all q_i and $r_j > 0$. Alternatively the propositions of section 2.1 can be suitably modified. In either case the bound given by Theorem 5 remains valid.

Before we provide the bound we need some further notation. Let

$$q_u = \underset{i \in I}{\text{Min}} \{q_i\}.$$

Let δ^j, $j \in J$ be the smallest value of δ such that

$$f_j(\delta^j) \geq K_0 - q_u - r_j. \tag{15}$$

If there exists no δ^j such that (15) is satisfied, let δ^j be the smallest value of δ that maximizes $f_j(\delta)$.

THEOREM 5: Let $\hat{B}$ be an optimal basis to $P3$ for all $\delta \geq \bar{\delta}$. Then $\hat{B}$ represents an optimal basis to $P3$ for $\delta \geq \hat{\delta}$ where $\hat{\delta} = \underset{j \in J}{\text{Max}}\, \delta^j$.

PROOF: From Proposition 1, the kth element of $\hat{B}^{-1}d(\delta)$ must be in one of the following forms:

$$\sum_{i \in I_k^+} q_i - \sum_{j \in J_k^-} r_j - \Big(\sum_{j \in J_k^-} f_j(\delta) \Big) \geq 0, \text{ or} \tag{16}$$

$$-\sum_{i \in I_k^-} q_i + \sum_{j \in J_k^+} r_j + \Big(\sum_{j \in J_k^+} f_j(\delta) \Big) \geq 0, \tag{17}$$

where I_k^+, I_k^-, J_k^+, and J_k^- are as defined in the proof of Theorem 2 and satisfy the conditions (a), (b), and (c) specified there.

If the kth element is given by (16), its value is ≥ 0 for all $\delta \geq 0$. This is because (16) holds for all $\delta \geq \bar{\delta}$ and $f_j(\delta)$ is a nondecreasing nonnegative function of δ. Now suppose the kth element is given by (17). We then have

$$\sum_{j \in J_k^+} f_j(\delta) \geq \sum_{i \in I_k^-} q_i - \sum_{j \in J_k^+} r_j \qquad \text{for all } \delta \geq \bar{\delta}. \tag{18}$$

If $I_k^- = I$, then $J_k^+ = J$ by condition (b) stated in the proof of Theorem 2. Thus an upper bound for the right hand side of (18) is given by $\text{Max} \left\{ 0, K_0 - q_u - \sum_{j \in J_k^+} r_j \right\}$. Furthermore $\hat{\delta}$ is such that for each $j \in J$ either $f_j(\hat{\delta}) \geq K_0 - q_u - r_j$ or $f_j(\hat{\delta})$ maximizes $f_j(\delta)$. Thus for each $j \in J_k^+$ we have either $f_j(\hat{\delta}) \geq K_0 - q_u - r_j \geq K_0 - q_u - \sum_{j \in J_k^+} r_j$ or $f_j(\hat{\delta}) \geq f_j(\bar{\delta})$. Consequently, since (18) is satisfied for all $\delta \geq \bar{\delta}$, it must be satisfied for $\delta \geq \hat{\delta}$. Thus $\hat{B}$ is feasible and hence optimal for all $\delta \geq \hat{\delta}$. This completes the proof of Theorem 5.

It is easily seen that for $\delta \geq \hat{\delta}$ in the network (V_3, E_3) the unique path (Γ_j) from s to j as defined by $\hat{B}$ is of the form

$$\Gamma_j = \{(s, i_j), (i_j, j)\} \text{ for } j \in J \text{ where } i_j \in I. \tag{19}$$

Theorem 5 is important, since for $\delta \geq \hat{\delta}$ the optimal growth path becomes invariant. Further from (19), in market j all increases in demand beyond $\hat{\delta}$ is supplied by a single source i_j through route (i_j, j).

$Z(\delta)$ is piecewise linear and convex over each interval $\delta_i \leq \delta \leq \delta_{i+1}$ $i=0, 1, \ldots, p$, but not necessarily convex over the entire range. Now suppose $f_j(\delta) = \beta_j \cdot \delta$ for all $j \epsilon J$ with $\beta_j \geq 0$. Although $Z(\delta)$ is now convex, we can easily construct examples in which the "more for less" paradox is obtained. When each β_j is a nonnegative integer the bound $\hat{\delta}$ for the prespecified market growth rate problem is less than or equal to the corresponding bound for the logistics growth problem. However, corresponding to this bound $\hat{\delta}$, the total shipment in the system is in general greater for this problem than the logistics growth problem.

4. CONCLUDING REMARKS

In sections 2 and 3, we have derived some characterizations of the optimal growth paths for a number of transportation-type problems which have been considered in the literature. This is achieved by an analysis of their respective network structures. These problems can also be converted into uncapacitated transportation problems (see [7] and [16]). Though such a conversion is desirable from a computational point of view, the resulting equivalent transportation networks do not easily lead to a succinct characterization of the optimal growth paths. The general spirit of the analysis of sections 2 and 3 can also be extended to the parametric transportation problem. This is discussed in detail in [7].

REFERENCES

[1] Balas, E. and P. Ivanescu, "On the Transportation Problem—I and II," Cahiers du Centre de Recherche Operationnelle *4*, 98–159 (1962).

[2] Charnes, A. and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming* (John Wiley and Sons, Inc., New York, New York, 1961), Vols. I and II.

[3] Charnes, A. and W. W. Cooper, "Systems Evaluation and Repricing Theorems," Management Science *9*, 33–49 (1962).

[4] Charnes, A., F. Glover, and D. Klingman, "A Note on a Distribution Problem," Operations Research *18*, 1213–1266 (1970).

[5] Charnes, A. and D. Klingman, "The More-for-Less Paradox in the Distribution Model," Cahiers du Centre d'Etudes de Recherche Operationnelle *13*, 11–22 (1971).

[6] Dantzig, G. B., *Linear Programming and Extensions* (Princeton University Press, Princeton, New Jersey, 1963).

[7] Fong, C. O. and M. R. Rao, "Parametric Studies in the Transportation and Transportation-Type Problems," Working Paper 7324, Graduate School of Management, University of Rochester, Rochester, New York, (Oct. 1973).

[8] Ford, L. R. and D. R. Fulkerson, *Flows in Networks* (Princeton University Press, Princeton, New Jersey, 1963).

[9] Grabowski, W. and Szwarc, W., "Optimum Transportation Bases" Lastosowania Mathematyki *9*, 357–389 (1968).

[10] Minieka, E., "A Note on Parametric Network Flows," Management Science *19*, 585–587 (1973).

[11] Sakarovitch, M., *Notes on Linear Programming* (Van Nostrand Reinhold Company, New York, New York, 1971).

[12] Simmonard, M., *Linear Programming* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1966).

[13] Srinivasan, V. and G. L. Thompson, "An Operator Theory of Parametric Programming for the Transportation Problem—I and II," Nav. Res. Log. Quart. *19*, 205–252 (1972).

[14] Srinivasan, V. and G. L. Thompson, "Determining Optimal Growth Paths in Logistics Operations," Nav. Res. Log. Quart. *19*, 575–599 (1972).

[15] Szwarc, W., "The Transportation Paradox," Nav. Res. Log. Quart. *18*, 185–202 (1971).

[16] Wagner, H. M., "The Lower Bounded and Partial Upper Bounded Distribution Model," Nav. Res. Log. Quart. *20*, 265–269 (1973).

ADJACENT VERTICES ON TRANSPORTATION POLYTOPES

Patrick G. McKeown
State University of New York at Albany
and
David S. Rubin
University of North Carolina at Chapel Hill

ABSTRACT

To solve linear fixed charge problems with Murty's vertex ranking algorithm, one uses a simplex algorithm and a procedure to determine the vertices adjacent to a given vertex. In solving fixed charge transportation problems, the simplex algorithm simplifies to the stepping-stone algorithm. To find adjacent vertices on transportation polytopes, we present a procedure which is a simplification of a more general procedure for arbitrary polytopes.

I. INTRODUCTION

Murty has suggested a vertex ranking procedure for solving the general linear fixed charge problem [8]. One of the crucial steps in implementing that algorithm is the determination of all vertices adjacent to a given vertex. When the feasible polyhedron is nondegenerate, the simplex method can be used to find these adjacent vertices. When degeneracy is present, however, the simplex algorithm cannot be used unless all basic representations of the given vertex are known. We have previously [6, 9] suggested the use of Chernikova's algorithm [3, 4] to find adjacent vertices when degeneracy occurs.

In his dissertation [6], McKeown has used Chernikova's algorithm to implement Murty's vertex ranking procedure for the general linear fixed charge problem when degeneracy is encountered. He also tested the general Chernikova procedure on some degenerate fixed charge transportation problems. He successfully solved some 4×6 problems (i.e., four supply points and six demand points), but was unable to solve 5×7 problems. For these larger problems a vast number of vertices had to be generated and stored, and he was forced to terminate prematurely due to time and storage considerations.

It seems reasonable to suppose that larger fixed charge transportation problems could be solved at reasonable computational cost if the procedures used to implement Murty's algorithm were specialized to take advantage of the structure of the transportation problem. The simplex algorithm can be replaced by the primal-dual algorithm or the stepping-stone algorithm, the latter being just a specialization of the simplex algorithm. In this paper, we show how to specialize Chernikova's algorithm for the transportation problem; a later paper will report computational results [7].

In section II we discuss the transportation polytope and some adjacency relations among its vertices. Section III presents Chernikova's procedure and shows how to specialize that procedure for transportation polytopes. An example is presented in the last section.

II. TRANSPORTATION POLYTOPES

Consider the transportation constraints

$$\sum_{j=1}^{n} x_{ij} = \alpha_i \qquad i = 1, \ldots, m$$

$$\sum_{i=1}^{m} x_{ij} = \beta_j \qquad j = 1, \ldots, n \tag{1}$$

$$x_{ij} \geq 0 \qquad i = 1, \ldots, m; \qquad j = 1, \ldots, n.$$

Here the α_i and β_j are positive real numbers (ordinarily integers) such that $\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{n} \beta_j$. The set of all x satisfying (1) is a bounded convex polyhedron; we shall refer to it as a *transportation polytope.* Previous studies of transportation polytopes [2, 5, 10] have been concerned with counting the number of facets and vertices of particular transportation polytopes and investigating the different combinatorial types of transportation polytopes for given m and n as the α_i and β_j vary. We have not discovered any previous discussion of vertex adjacency.

It will be convenient to write (1) in matrix form as $\bar{A}x = \bar{b}$, $x \geq 0$. The matrix $\bar{A}$ is $(m+n) \times mn$. It is the node-arc incidence matrix of the transportation network, and hence is totally unimodular with rank $m+n-1$. Since any $m+n-1$ rows of $\bar{A}$ are linearly independent, the last constraint of (1) is redundant. We shall delete that last constraint, and work with the equivalent system $Ax = b$, $x \geq 0$. Here A is $\bar{A}$ less its last row. We are concerned with $T = \{x | Ax = b, x \geq 0\}$.

If B is any $(m+n-1) \times (m+n-1)$ nonsingular submatrix of A, we say that B is a *basis* for A; we partition A as (B, N), by rearranging columns if necessary, and similarly partition x as $\begin{pmatrix} x_B \\ x_N \end{pmatrix}$. We call the variables x_B *basic*, and the variables x_N *nonbasic*. The basis B is *feasible* if $B^{-1}b \geq 0$, and then the point $\bar{x} = \begin{pmatrix} \bar{x}_B \\ \bar{x}_N \end{pmatrix} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ is a *basic feasible solution* of (1). It is well known that all basic feasible solutions of (1) are vertices of T and conversely.

If $\bar{x}_B > 0$, we say that the basic feasible solution $\bar{x}$ is *nondegenerate.* If all basic feasible solutions of (1) are nondegenerate, then we say that T is nondegenerate. If T is nondegenerate, then there is a one-to-one correspondence between vertices of T and feasible bases for A; if T is degenerate, several bases may yield the same vertex. We say that two feasible bases are *algebraically adjacent* if they differ by one column. Two vertices of T are *geometrically adjacent* if there exists a supporting hyperplane to T whose intersection with T is the line segment joining the two vertices. The line segment joining two adjacent vertices of T is an *edge* of T. When T is nondegenerate, the notions of algebraic and geometric adjacency are equivalent, i.e., x^1 and x^2 are geometrically adjacent vertices if and only if the corresponding bases are algebraically adjacent. For degenerate T, the situation is more complex, but may be summed up as follows:

1. If B^1 and B^2 are algebraically adjacent feasible bases, with corresponding vertices x^1 and x^2, then either $x^1 = x^2$ or else x^1 and x^2 are geometrically adjacent vertices.

2. If x^1 and x^2 are geometrically adjacent vertices, then there exist corresponding bases B^1 and B^2 which are algebraically adjacent.

We are interested in geometric adjacency, but since the simplex algorithm is to be discussed, we must deal with algebraic adjacency, because a single simplex pivot moves from one basis to an adjacent basis; only if the pivot ratio is positive does the pivot move to an adjacent vertex.

Relative to a given feasible basis, we may represent the constraints $Ax=b$, $x\geq 0$ by Tucker's tableau

$$x_B=B^{-1}b+B^{-1}N(-x_N)\geq 0$$

$$x_N=0+(-I)(-x_N)\geq 0,$$

which we abbreviate as

$$x=y_0-Yx_N\geq 0.$$

Here y_0 is $mn+1$ and Y is $mn\times(mn-m-n+1)$. We let y_j denote the jth column of Y.

Since $\bar{A}$ is totally unimodular, all the elements of Y are 0, +1, or −1. Let $\theta_j=\min\{y_{i0}|y_{ij}=1\}$. If $\theta_j>0$, then the line segment $\{x=y_0-\theta y_j|0\leq\theta\leq\theta_j\}$ is an edge of T. Then from point 2 above we have

LEMMA 1: If x^1 and x^2 are adjacent vertices of T, then $x^2-x^1=-\theta y$, where θ is a scalar and y is a vector with components 0, +1, or −1.

We shall need a stronger characterization of these "edge directions" y. For this purpose, we introduce a concept of "decomposability" similar to that used by Balas and Padberg [1] in discussing the set partitioning problem. Let x^1 and x^2 be two vertices of T such that $-\theta y=x^2-x^1$ has all elements $0, +\theta$, or $-\theta$. We say that y is *decomposable* if there exist noncollinear $\bar{y}$ and $\bar{\bar{y}}$, such that $y=\bar{y}+\bar{\bar{y}}$, and both $\bar{x}=x^1-\theta\bar{y}$ and $\bar{\bar{x}}=x^1-\theta\bar{\bar{y}}$ are in T.

LEMMA 2: Let x^1 and x^2 be vertices of T. If x^1 and x^2 are adjacent, then $y=-(x^2-x^1)/\theta$ is not decomposable.

PROOF: Let $L=\{x=\lambda x^1+(1-\lambda)x^2|0\leq\lambda\leq 1\}$. Let $H=\{x|h^Tx=\eta\}$ be a supporting hyperplane to T such that $L=H\cap T$ and for $x\epsilon T$ we have $h^Tx\leq\eta$. Such hyperplanes exist since L is an edge of T. Suppose that y is decomposable. Note that x^1, x^2, and $\bar{x}$ are not collinear since $\bar{y}$ and $\bar{\bar{y}}$ are not collinear. Now

$$x^1, x^2\epsilon L \Rightarrow \eta=h^Tx^1=h^Tx^2=h^T(x^1-\theta y)=\eta-\theta h^Ty \Rightarrow h^Ty=0.$$

Also

$$\bar{x}\epsilon T \Rightarrow \eta\geq h^T\bar{x}=h^T(x^1-\theta\bar{y})=\eta-\theta h^T\bar{y} \Rightarrow h^T\bar{y}\geq 0.$$

Similarly we have $h^T\bar{\bar{y}}\geq 0$.

$$\therefore 0=h^Ty=h^T(\bar{y}+\bar{\bar{y}})\geq 0$$

$$\therefore h^T\bar{y}=h^T\bar{\bar{y}}=0.$$

Hence $\bar{x}$ and $\bar{\bar{x}}$ lie in $H \cap T = L$. But this contradicts our remark that x^1, x^2, and $\bar{x}$ are not collinear. Hence y cannot be decomposable. Q.E.D.

In addition to Tucker's tableau, it will also be convenient to use the transportation tableau $\mathscr{T}$, an $m \times n$ array whose cell in the ith row and jth column represents the variable x_{ij}. We use the term *line of* $\mathscr{T}$ to mean either a row or column of $\mathscr{T}$.

THEOREM 1. Let x^1 and x^2 be two geometrically adjacent vertices of T. Then no line of $\mathscr{T}$ contains two cells whose associated variables have larger values at x^2 than at x^1.

PROOF: Let $x^2 - x^1 = -\theta y$. If the ijth component of y is $+1$ (-1), mark cell (i, j) with a $-$ $(+)$. The marks indicate how x_{ij} changes in moving from x^1 to x^2. From Lemma 1 and the constraints (1), it follows that in any line of $\mathscr{T}$, the number of $+$ marks equals the number of $-$ marks.

Now suppose that some line of $\mathscr{T}$, say row i_1, contains two or more $+$ marks. Suppose the marks in question are in columns j_0 and j_1. Column j_1 contains a $-$ mark in row $i_2 \neq i_1$; row i_2 contains a $+$ mark in column $j_2 \neq j_1$; and so on. Thus we trace out a sequence of cells and marks

$$\begin{matrix}(i_1, j_1), & (i_2, j_1), & (i_2, j_2), & \ldots, & (i_k, j_k), & (i_{k+1}, j_k) & \ldots . \\ + & - & + & & + & - & \end{matrix}$$

Since $\mathscr{T}$ contains only a finite number of lines, eventually we get some $i_k = i_{k-l}$ for some $l \geq 1$ or else $j_k = j_{k-l}$ for some $l \geq 1$. Without loss of generality, we may assume that the first such index to repeat is $i_k = i_1$, so the sequence is

$$\begin{matrix}(i_1, j_1), & (i_2, j_1), & (i_2, j_2), & \ldots, & (i_k, j_{k-1}) = (i_1, j_{k-1}). \\ + & - & + & & - & \end{matrix}$$

The sequence is a familiar "rook's tour", and from it we construct $\bar{y}$ as follows

$$\bar{y}_{ij} = \begin{cases} +1 \text{ if } (i, j) \text{ is on the tour and marked } - \\ -1 \text{ if } (i, j) \text{ is on the tour and marked } + \\ 0 \text{ if } (i, j) \text{ is not on the tour.} \end{cases}$$

Let $\bar{\bar{y}} = y - \bar{y}$. Since (i_1, j_{k-1}) is marked $-$, $j_{k-1} \neq j_0$, and hence $\bar{\bar{y}} \neq 0$ (in particular, $\bar{\bar{y}}_{i, j_0} = -1$.) Clearly $\bar{y}$ and $\bar{\bar{y}}$ are not collinear, and $\bar{x} = x^1 - \theta y$ and $\bar{\bar{x}} = x^1 - \theta y$ are both in T. Thus y is decomposable, which contradicts Lemma 2, and hence proves the theorem. Q.E.D.

COROLLARY: Let x^1 and x^2 be geometrically adjacent vertices of T. Let B^1 be a basis corresponding to x^1. Then the number of variables which are nonbasic in B and positive at x^2 is at most $\min\{m, n\}$.

REMARK: If the "larger" in the statement of the theorem is replaced by "smaller," the result is still true.

III. CHERNIKOVA'S ALGORITHM

Let M be a real $p \times q$ matrix. We wish to find all the edges (extreme rays) of the cone $C = \{w \mid Mw \geq 0, w \geq 0\}$. Chernikova has presented a procedure for doing this [3, 4]. We present a paraphrase of that procedure due to Rubin [9].

Consider the matrix $\begin{pmatrix} -M \\ -I \end{pmatrix}$, where I is a $q \times q$ identity matrix. Chernikova's algorithm gives a series of transformations of this matrix which generates all the edges. At any stage of the process we denote the old matrix by $Y = \begin{pmatrix} U \\ L \end{pmatrix}$, and the new matrix being generated is denoted $\bar{Y}$. The matrices U and L will always have p and q rows, respectively; however, they will not have q columns. They will have more than q columns in most cases, but if C lies in some subspace of R^q they may have fewer than q columns. For $w \in R^q$, we use the symbol (w) to denote the ray $\{\lambda w, \lambda \geq 0\}$.

The algorithm is as follows:

(0.0) If any row of U has all components positive, then $w = 0$ is the only point in C.

(0.1) If all the elements of U are nonpositive, then the negatives of the columns of L are the edges of C, i.e., the ray $(-l_j) = \{w = -\lambda l_j, \lambda \geq 0\}$ is an edge of C; here l_j denotes the jth column of L.

(1) Choose the first row of U, say row r, with at least one positive element.

(2) Let $R = \{j \mid y_{rj} \leq 0\}$. Let $v = |R|$, i.e., the number of elements of R. Then the first v columns of the new matrix, $\bar{Y}$, are all the y_j for $j \in R$, where y_j denotes the jth column of Y.

(2'.) If Y has only two columns and $y_{r1}y_{r2} < 0$, adjoin the column $|y_{r2}|y_1 + |y_{r1}|y_2$ to the $\bar{Y}$ matrix. Go to step 4.

(3) Let $S = \{(s, t) \mid y_{rs}y_{rt} < 0, s < t\}$, i.e., the set of all (unordered) pairs of columns of Y whose elements in row r have opposite signs. Let I_0 be the index set of all nonpositive rows of Y. For each $(s, t) \in S$, find all $i \in I_0$ such that $y_{is} = y_{it} = 0$. Call this set $I_1(s, t)$. We now use some of the elements of S to create additional columns for $\bar{Y}$:

(a) If $I_1(s, t) = \phi$ (the empty set), then y_s and y_t do not contribute another column to the new matrix.

(b) If $I_1(s, t) \neq \phi$, check to see if there is a u not equal to either s or t, such that $y_{iu} = 0$ for all $i \in I_1(s, t)$. If such a u exists, then y_s and y_t do not contribute another column to the new matrix. If no such u exists, then choose $\alpha_1, \alpha_2 > 0$ to satisfy $\alpha_1 y_{rs} + \alpha_2 y_{rt} = 0$. (We always choose $\alpha_1 = |y_{rt}|, \alpha_2 = |y_{rs}|$.) Adjoin the column $\alpha_1 y_r + \alpha_2 y_s$ to the new matrix.

(4) When all pairs in S have been examined, and the additional columns (if any) have been added, we say that row r has been "processed." Now let Y denote the matrix $\bar{Y}$ produced in processing row r, and return to step 0.0.

A proof of the algorithm may be found in References [3] and [4]. A geometric discussion may be found in [9].

Let us now consider the simplex tableau for the transportation problem

$$x_B = B^{-1}b + B^{-1}N(-x_N) \geq 0$$

$$x_N = 0 + (-I)(-x_N) \geq 0.$$

Suppose that the degenerate basic variables (those which are zero) are the last components of x_B. We may write the tableau as

$$x_Q = q + \quad Q(-x_N) \geq 0$$

$$x_D = 0 + \quad D(-x_N) \geq 0$$

$$x_N = 0 + (-I)(-x_N) \geq 0,$$

where $q > 0$. Note that x_D has at most $\min\{m, n\} - 1$ components since all α_i and β_j are positive and hence each line of $\mathscr{T}$ must contain at least one positive component of x_B. Finding the vertices of T adjacent to $x^1 = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$ is equivalent to finding the edges of the cone $C = \{x_N \mid -Dx_N \geq 0,\ x_N \geq 0\}$. (for details of this, see [9].) Thus the simplex tableau less the nondegenerate rows can serve as the initial Y matrix for Chernikova's algorithm. If the final tableau in the algorithm is $\begin{pmatrix} U \\ L \end{pmatrix}$, then the rows corresponding to the nondegenerate basic variables have entries $-QL$.

LEMMA 3: At any iteration of Chernikova's algorithm, all elements of U, L, *and* $-QL$ are $0, +1$, or -1.

PROOF: We proceed inductively. Suppose D has only one row. When two columns are combined the "weights" are $\alpha_1 = \alpha_2 = 1$, and the columns produced by the algorithm correspond to edges of T, and the result follows from Lemma 1. Suppose the result is true if D has k rows. Consider a D with $k+1$ rows. Let x_{ij} be the last component of x_D. If we perturb the original data by increasing α_i and β_j by one, the initial tableau corresponding to the same basis will be identical to the tableau under discussion, except that x_{ij} will have the value of $+1$ at the vertex in question. The other rows of the tableau would be unchanged. Hence by the induction hypothesis, all columns produced in processing rows 1 through k of D will have elements of $0, +1$, or -1. Then Lemma 1 guarantees that the same will be true for those columns produced by processing row k, so the result follows by induction. Note also that we always have $\alpha_1 = \alpha_2 = 1$. Q.E.D.

This result has computer time and storage implications which are discussed at length in [7]. We now show how to simplify Chernikova's algorithm for the transportation problem.

Given a basis for the problem, the Srinivasan-Thompson procedure [11] can be used to determine the components of $B^{-1}N$ and $B^{-1}b$. From this we can obtain D. We refer to columns in the tableau by two systems: a single subscript k indicates that we refer to the kth column in the tableau; a double subscript ij refers to the variable associated with the column. For initial columns k or ij (before Chernikova's procedure is begun), let $R(k) = \{i\}$, $C(k) = \{j\}$. Then when columns s and t are combined by the algorithm to get column r, let

$$R(r) = R(s) \cup R(t),\ C(r) = C(s) \cup C(t).$$

The following result is a restatement of Theorem 1:

THEOREM 2: If $R(s) \cap R(t) \neq \phi$ (the empty set) or $C(s) \cap C(t) \neq \phi$, then columns s and t should not be combined.

This is clearly stronger than Chernikova's zero checking since it requires only $m+n$ rows of information (which can be stored as bits) rather than the $(m-1)(n-1)$ rows corresponding to x_N plus those corresponding to x_D.

If $R(s) \cap R(t) = C(s) \cap C(t) = \phi$, then we must consider Chernikova's zero checking more closely. For original columns k or ij let $I(k) = \{ij\}$. When columns s and t are combined to get column r, let

$$I(r) = I(s) \cup I(t).$$

From the Corollary to Theorem 1, it follows that $I(r)$ has at most min $\{m, n\}$ elements. The following result is straightforward.

LEMMA 4: If $I(u) \not\subseteq I(s) \cup I(t)$, then $y_{iu} \neq 0$ for all $i \in I_1(s, t)$. ($I_1(s, t)$ is defined in step 3 of the algorithm.)

Theorem 2 and Lemma 4 allow us to simplify step 3 of the algorithm as follows:

(3*) Let $S = \{(s, t) \mid y_{rs} y_{rt} < 0, s < t\}$, I_0 be the index set of all nonpositive rows of U. For each $(s, t) \in S$, let $I_1(s, t) = \{i \in I_0 \mid y_{is} = y_{it} = 0\}$. We now use some of the elements of S to create additional columns for Y.

(a) If $R(s) \cap R(t) \neq \phi$ or $C(s) \cap C(t) \neq \phi$, don't combine columns s and t. Get the next element of S, and go to (a).

(b) For each column $u \neq s, t$, proceed as follows:

(i) If $I(u) \not\subseteq I(s) \cup I(t)$, go to the next u.

(ii) If $I(u) \subseteq I(s) \cup I(t)$, check to see if $y_{iu} = 0$ for all $i \in I_1(s, t)$ or if $I_1(s, t) = \phi$. If so, don't combine s and t, get the next element of S, and go to (a).

(iii) If all $u \neq s, t$ have been checked and none prevented the combination of s and t, then add the column $y_s + y_t$ to $\bar{Y}$, get the next element of S, and go to (a).

Replacing the matrix L with the index sets R, C, and I, and replacing step 3 with step 3* gives the simplified algorithm for neighboring vertices on transportation polytopes.

IV. AN EXAMPLE

Consider the 4×4 transportation problem given by the following $\mathcal{T}$

		0	3
	0	4	
0	8		
6			

The basic variables and their values are entered in $\mathcal{T}$; cells with no entries are nonbasic. For this basis, the condensed (i.e., without the rows for x_N) Tucker's tableau is:

		$-x_{11}$	$-x_{12}$	$-x_{21}$	$-x_{24}$	$-x_{33}$	$-x_{34}$	$-x_{42}$	$-x_{43}$	$-x_{44}$
x_Q	$x_{14}=3$	0	0	0	1	0	1	0	0	1
	$x_{23}=4$	−1	−1	0	1	1	1	0	1	1
	$x_{32}=8$	−1	0	−1	0	1	1	1	1	1
	$x_{41}=6$	0	0	0	0	0	0	1	1	1
x_D	$x_{13}=0$	1	1	0	−1	0	−1	0	0	−1
	$x_{22}=0$	1	1	1	0	−1	−1	0	−1	−1
	$x_{31}=0$	1	0	1	0	0	0	−1	−1	−1

From this we may obtain the initial Chernikova tableau by dropping the nondegenerate rows:

	1	2	3	4	5	6	7	8	9
x_{13}	1	1	0	−1	0	−1	0	0	−1
x_{22}	1	1	1	0	−1	−1	0	−1	−1
x_{31}	1	0	1	0	0	0	−1	−1	−1
R	1	1	2	2	3	3	4	4	4
C	1	2	1	4	3	4	2	3	4
I	11	12	21	24	33	34	42	43	44

In processing the x_{13} row, all possible combinations pass the tests of step (3*), and so are made, yielding the second tableau (the columns after the vertical bar were created during the iteration):

	1	2	3	4	5	6	7	8	9	10	11	12	13
x_{13}	0	−1	0	−1	0	0	−1	0	0	0	0	0	0
x_{22}	1	0	−1	−1	0	−1	−1	1	0	0	1	0	0
x_{31}	1	0	0	0	−1	−1	−1	1	1	0	0	0	−1
R	2	2	3	3	4	4	4	1	1	1	1	1	1
								2	3	4	2	3	4
C	1	4	3	4	2	3	4	1	1	1	2	2	2
								4	4	4	4	4	4
I	21	24	33	34	42	43	44	11	11	11	12	12	12
								24	34	44	24	34	44

We now process the x_{22} row. The combination of column 1 with columns 3, 4, 6, or 7 passes all the tests. When we try to combine column 8 with column 3, the test of step (3*(a)) is passed since $R(3) \cap R(8) = \{3\} \cap \{1, 2\} = \phi$, $C(8) = \{3\} \cap \{1, 4\} = \phi$. $I(3) \cup I(8) = \{33, 11, 24\}$; note that $I(2) \subseteq \{33, 11, 24\}$, but columns 3 and 8 have their only common 0 in the x_{13} row, and column 2 does not have a 0 in that row; hence the test of step (3*(b)) is passed, and so columns 3 and 8 are combined. Columns 4 and 8 fail the step (3*(a)) test because $C(4) \cap C(8) = \{4\} \cap \{1, 4\} = \{4\} \neq \phi$; columns 6 and 8 pass all tests; columns 7 and 8 fail the step (3*(a)) test. Columns 3 and 11 and columns 6 and 11

pass all tests; the other possible combinations with column 11 (i.e., columns 4 and 7) fail the step (3*(a)) test. Hence we get the third tableau:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x_{13}	−1	0	−1	0	0	−1	0	0	0	0	0	−1	0	−1	0	0	0	0
x_{22}	0	−1	−1	0	−1	−1	0	0	0	0	0	0	0	0	0	0	0	0
x_{31}	0	0	0	−1	−1	−1	1	0	0	−1	1	1	0	0	1	0	0	−1
R	2	3	3	4	4	4	1	1	1	1	2	2	2	2	1	1	1	1
							3	4	3	4	3	3	4	4	2	2	2	2
															3	4	3	4
C	4	3	4	2	3	4	1	1	2	2	1	1	1	1	1	1	2	2
							4	4	4	4	3	4	3	4	3	3	3	3
															4	4	4	4
I	24	33	34	42	43	44	11	11	12	12	21	21	21	21	11	11	12	12
							34	44	34	44	33	34	43	44	24	24	24	24
															33	43	33	43

We now process the x_{31} row. Comments about possible combinations follow:

7 and 4, 7 and 5—all tests passed

7 and 6, 7 and 10, 7 and 18—(3*(a)) test failed

11 and 4, 11 and 10—all tests passed

11 and 5, 11 and 18—(3*(a)) test failed

11 and 6—(3*(a)) test passed; $I(6) \cup I(11) = \{21, 33, 44\}$, but $I(2) = \{33\} \subseteq \{21, 33, 44\}$ and $I_1(6, 11) = \phi$, so the step (3*(b)) test is failed. Note that this shows that the converse of Theorem 2 is false.

12 and 4—all tests passed

12 and 5—(3*(a)) test passed, (3*(b)) test failed ($I_1(5, 12) = \phi$)

12 and 6, 12 and 10, 12 and 18—(3*(a)) test failed

15 and 4—all tests passed

15 and 5, 6, 10 or 18—test (3*(a)) failed.

Thus the final tableau is (we suppress R and C since they are no longer needed, and add on the rows corresponding to x_Q):

						*								*		*		*		*
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$x_{14}=3$	1	0	1	0	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1
$x_{23}=4$	1	1	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1	1	1
$x_{32}=8$	0	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1
$x_{41}=6$	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1
$x_{13}=0$	−1	0	−1	0	0	−1	0	0	0	0	−1	0	0	0	0	0	0	0	−1	0
$x_{22}=0$	0	−1	−1	0	−1	−1	0	0	0	0	0	0	0	0	0	−1	0	0	0	0
$x_{31}=0$	0	0	0	−1	−1	−1	0	0	−1	0	0	0	0	−1	0	0	0	0	0	0
I	24	33	34	42	43	44	11	12	12	21	21	11	12	12	11	11	21	12	21	11
							44	34	44	43	44	24	24	24	34	34	33	21	34	24
												43	33	43	42	43	42	33	42	33
																		44		42

It should be remarked that it is only a coincidence that all elements in the x_Q rows are 0 or +1; in general some −1 elements can be found there. From this tableau we can easily find all 20 of the adjacent vertices. We illustrate the process with a few examples:

Column	θ	*Neighbor*
1	3	$x_{23}=1,\ x_{32}=8,\ x_{41}=6,\ x_{13}=3,\ x_{24}=3$
2	4	$x_{14}=3,\ x_{32}=4,\ x_{41}=6,\ x_{22}=4,\ x_{33}=4$
10	4	$x_{14}=3,\ x_{32}=8,\ x_{41}=2,\ x_{21}=4,\ x_{43}=4$
18	3	$x_{23}=1,\ x_{32}=5,\ x_{41}=3,\ x_{12}=3,\ x_{21}=3,\ x_{33}=3,\ x_{44}=3.$

We note that of the 20 adjacent vertices, only the five obtained from the starred columns are non-degenerate. In terms of degeneracy, transportation polytopes seem to be like the little girl with the little curl.

REFERENCES

[1] Balas, E. and M. Padberg, "On the Set Covering Problem, II: An Algorithm," Management Sciences Research Report No. 295, Carnegie-Mellon University (Nov. 1972).

[2] Bolker, E. D., "Transportation Polytopes," Journal of Combinatorial Theory (Series B) *13*, 251–262 (1972).

[3] Chernikova, N. V., "Algorithm for Finding a General Formula for the Nonnegative Solutions of a System of Linear Equations," U.S.S.R. Computational Mathematics and Mathematical Physics *4*, 151–158 (1964).

[4] Chernikova, N. V., "Algorithm for Finding a General Formula for the Nonnegative Solutions of a System of Linear Inequalities," U.S.S.R. Computational Mathematics and Mathematical Physics *5*, 228–233 (1965).

[5] Klee, V. and C. Witzgall, "Facets and Vertices of Transportation Polytopes," in G. B. Dantzig and A. F. Veinott, Jr. (eds), *Mathematics of the Decision Sciences* (Part 1), Lectures in Applied Mathematics, XI, (American Mathematical Society, Providence, Rhode Island, 1968) 257–282.

[6] McKeown, P. G., "An Extreme Point Ranking Algorithm for the Linear Fixed Charge Problem," unpublished doctoral dissertation, University of North Carolina at Chapel Hill (May 1973).

[7] McKeown, P. G. and D. S. Rubin, "An Extreme Point Ranking Algorithm for the Fixed Charge Transportation Problem," in preparation.

[8] Murty, K. G., "Solving the Fixed Charge Problem by Ranking the Extreme Points," Operations Research *16*, 268–279 (1968).

[9] Rubin, D. S., "Neighboring Vertices on Convex Polyhedral Sets," University of North Carolina (Aug. 1972).

[10] Simonnard, M. A. and G. F. Hadley, "Maximum Number of Iterations in the Transportation Problem," Nav. Res. Log. Quart. *6*, 125–129 (1959).

[11] Srinivasan, V. and G. L. Thompson, "Accelerated Algorithms for Labelling and Relabelling of Trees, with Applications to Distribution Problems," Journal of the ACM *20*, 712–726 (1972).

THE STATIC THEORY OF TRANSFER PRICING

Hermann Enzer
Department of the Interior
Washington, D.C.

ABSTRACT

An analysis of the literature of transfer pricing is presented. It is shown that, under assumptions that the firm and its divisions have full deterministic knowledge of their costs and demands, some form of average cost is the appropriate transfer price. What happens when a firm adopts an objective other than profit maximization is further examined.

1. INTRODUCTION

It has been claimed that four-fifths of all major U.S. corporations are largely or wholly decentralized [2]. If true this implies that most top managers must believe that the advantages of decentralization exceed the advantages of centralization in large corporations. There is little doubt that when decentralization is established with care, commitment, and knowledge, it is easier to evaluate the relative performance of a division and its manager in terms of economic criteria than under centralization. One of the ingredients of a successfully decentralized firm is a system of transfer prices at which products are transferred between the divisions of the firm that guides the divisions to accomplish the objectives of the firm through suboptimization. Since we know that suboptimizations by individual divisions of a firm do not necessarily result in an optimum for the firm as a whole, the transfer pricing problem consists in fixing these prices in such a way that divisional suboptimizations imply an overall optimum. Several authors [3, 9] have stated that this problem is inherently insoluble because firms have multiplicities of objectives and it is not possible to design a system of transfer prices which attains *all* objectives of a firm optimally. This may very well be the case. At any rate, although transfer prices are only one ingredient among many in decentralization, Adam Smith's invisible hand will not guide the parts toward the optimum of the whole unless considerable care is exercised in setting prices.

A variety of pricing policies are followed in practice. Products are transferred at average cost, average variable cost, variable cost plus a margin, marginal cost, market price, market price minus a margin, "long-run value," market-based negotiated price, and some intermediate forms adapted to particular situations. The theoretical basis of transfer pricing was provided by Hirshleifer [12, 13] who showed in two articles that under static conditions and under certainty the generally correct transfer price for a product is the marginal cost of the producing division. Hirshleifer made the usual simplifying assumptions, one of them being that the firm pursues the single objective of profit maximization. His solution has been accepted as the theoretically correct solution by apparently all authors on transfer pricing, and has become the basis of further extensions and elaborations. It is the purpose of this article to show that Hirshleifer's solution is incorrect. The correct generally valid transfer price under the same conditions as assumed by Hirshleifer is not marginal cost, but a form of average cost. Although transfers at marginal cost if possible will bring about correspondence between the suboptima

of the divisons and the optimum of the firm, they are neither theoretically correct nor very desirable as pricing rules. Aside from the usual problem of determining marginal costs in practice, when the optimum point lies on the increasing branch of a marginal cost curve the buying division will consider the transfer of the product at marginal cost as inequitable. And when the point lies on the decreasing branch, the selling division will be unhappy. Moreover, as we shall see, marginal cost pricing has to be accompanied by a stricture, already mentioned by Hirshleifer, which practically means that the buying division is told what quantity it should buy from the selling division. But this largely defeats the purpose of decentralization.

We will show in the following sections that for the typical cases discussed by Hirshleifer and others, some form of average cost is the appropriate transfer price. When constraints are included, the average cost has to be adjusted in terms of opportunity profits. We will further examine what happens when a firm adopts an objective other than profit maximization. It is assumed that the firm and its divisions have full deterministic knowledge of their costs and demands.

2. THE HIRSHLEIFER ANALYSIS

The following assumptions are made. A firm establishes two profit centers; i.e., the objective of each center is to maximize its respective profit. The objective of the firm as a whole is also profit maximization. The first division, to be called Division 1, is a manufacturing division whose output is transferred to Division 2, a distribution division, which finishes the product and sells it in the market. We shall confine ourselves in this section to the case when the output of Div. 1 does not have a market price. In other words, Div. 1 produces an intermediate product which is not traded in the market. Division 2 finishes the intermediate product and sells the final product at the prevailing market price, which we suppose is constant and given to the firm. Since each division operates as a profit center, the output of Div. 1 will be charged to Div. 2 and the problem of the firm is to fix a transfer price so that the independent actions of the two divisions will maximize the overall profit of the firm.

To simplify the analysis, we assume that there exists no technological dependence between the two divisions in the sense that the costs of one division are independent of the activity levels of the other divsion except for the costs of the intermediate product. This means that there is no joint-product relationship between the two divisions in the technical sense. Since our approach is static, inventory accumulation, depreciation of equipment, and other dynamic aspects can be ignored. To determine the output level of Div. 2 which maximizes the profit of the firm (remember, only Div. 2 operates in the market), we perform the usual operation of finding that output at which the marginal cost of the firm, MC, equals the market price, p,

$$MC = p.$$

The marginal cost of the firm is the sum of the marginal costs of the divisions,

$$MC = MC_1 + nMC_2,$$

where nMC_2 (net marginal cost) excludes the transfer price of the intermediate product. If we assume that the units of the intermediate and final products are commensurate in the sense that both have a specific quantitative relationship to each other, the equation

$$p = MC_1 + nMC_2$$

can then be solved for the optimum output levels.

Hirshleifer solves the problem of optimum outputs and the appropriate transfer price with the help of a diagram, Figure 1. Both the output of Div. 1, designated by u, and the output of Div. 2, designated by q, are measured along the same axis since commensurability between the two outputs is assumed. The optimum output for Div. 2, denoted by q^0, occurs where the rising marginal cost curve of the firm intersects the price line p. This output level can be reached through the suboptimizations of the divisions as follows. Div. 2 obtains from the manufacturing division a schedule showing what quantity of the intermediate product the manufacturing division would produce—and sell—at any transfer price r_u. This schedule would be the same as the MC_1 curve if the manufacturing division rationally determines its output and sets $MC_1 = r_u$. With this information the distribution division can derive a curve, labelled $p - r_u$ in the diagram, which represents the difference between market price and transfer price for any level of output. The distribution division then finds its output level by operating at the point where the "net marginal revenue" curve $p - r_u$ intersects nMC_2. The manufacturing division also produces at the same level because $MC_1 = r_u$ there. Hirshleifer then concludes that the transfer price should equal the marginal cost of the manufacturing division for the level q^0.

This solution causes several problems. First, for the case of Fig. 1, Div. 1 makes a "profit" represented by the area between the curve MC_1 and the line $r_u^0 r_u^0$. "Profit" is made whenever the solution lies on the rising branch of the marginal cost curve since the average cost curve is below the marginal curve. What happens, however, when the solution occurs on the downward sloping branch of the marginal cost curve as is frequently the case? The average cost curve then lies above the marginal curve and by setting the transfer price equal to marginal cost, Div. 1 would record a "loss." A transfer pricing system that is not equitable in the eyes of both divisions would have little chance accomplishing its goals. Second, the marginal cost pricing rule must be accompanied by a stipulation.[1] Division 2 must not be allowed to increase its profit by operating on a curve marginal to the curve $p - r_u$ (usually called a quasi-marginal curve in economics), the curve labelled mr in Fig. 1. If Div. 2 were permitted to move to the point where $mr = nMC_2$, it would sell the quantity q^* instead of the optimal q^0, and thereby increase the profit of Div. 2, but decrease the overall profit of the firm.[2] Division 2's profit for the output q^0 is represented by the area between the line segment ss and the nMC_2 curve. For the output level q^* the profit becomes the area between the line segment tt and the corresponding segment of nMC_2.

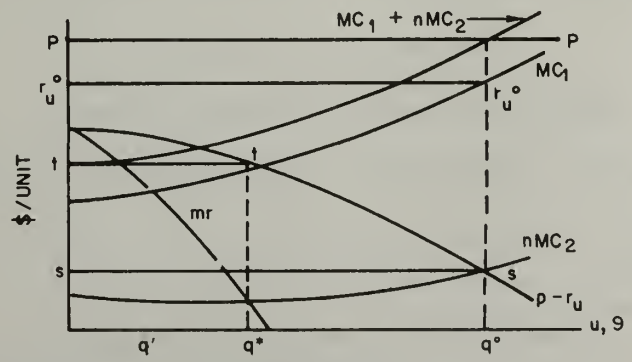


FIGURE 1

Since Div. 2's profit is larger at level q^* than at q^0, and since the curves are all continuous, Div. 2 can increase its profit by choosing any output level in the interval q^*q^0. The marginal cost rule must therefore be accompanied by the requirement that Div. 2 cannot operate in the interval q^*q^0, because any move away from q^0 toward q^* will reduce the profit of the firm. But this is not enough. A certain output interval to the left of q^*, say $q'q^*$, must also be declared off limits to Div. 2 since there exists a neighborhood to the left of q^* in which profits to Div. 2 are greater than at q^0. (Again, with continuous curves profit rises and declines gradually.) It follows that after we eliminate the interval $q'q^*$, Div. 2 as a profit maximizer is indifferent between outputs q' and q^0, and to make sure that the global optimum level q^0 is selected, we also stipulate that Div. 2 may not operate in the interval $0q'$. But this in effect means that we are telling Div. 2 to sell the quantity q^0 and the purpose of decentralized decision making has been largely defeated.

Third, for such decisions as whether or not to shut down a division, marginal rules which by definition represent values of incremental or decremental changes are of little use and one has to go back to total magnitudes.

Although practice has shown that there does not seem to exist a transfer price which cannot be faulted in some way, the relatively serious first and second drawbacks mentioned should have suggested that there may be something wrong with the marginal cost pricing rule.[3] We will now work out the correct transfer price under the same assumptions as for the Hirshleifer analysis.

3. NO MARKET FOR THE INTERMEDIATE PRODUCT

We shall employ a few relatively simple mathematical relations and manipulations. Assume that Div. 1 uses two inputs, x_1 and x_2, to produce the intermediate output u, and that inputs and outputs are related by an economic production function,

$$u=f(x_1, x_2).$$

This function summarizes (and suppresses many details important in real production) the generally rather complex production relationships in terms of the three stated variables. We shall suppose that this function and all other mathematical terms to be introduced are continuous and can be differentiated. Let us further designate the unit cost of input i by r_i $(i=1, 2)$ and treat it as a constant. The as yet unknown transfer price of the intermediate product will be denoted by r_u. Then if Div. 1 is a profit center, its problem is:

$$\text{maximize } \pi_1=r_uu-r_1x_1-r_2x_2$$

$$\text{subject to the production function } u=f(x_1, x_2),$$

where r_uu represents the receipts of the division, $r_1x_1+r_2x_2=C_1$ its costs, and the symbol π profits.

---

[1] Hirshleifer (1956, p. 175)

[2] The reader who wishes to see a numerical demonstration of the increase in profit, may consult Naert and Janssen who have worked out a simple example.

[3] The disadvantages of the marginal cost pricing rule have been recognized by many authors, Hirshleifer himself being one of the first. Particularly, Ronen and McKinney, Shillinglaw, and Solomons have discussed marginal cost pricing in detail and shown its defects.

We may adopt similar relations for Division 2. Suppose that Division 2 also uses two inputs, x_3 and u, to produce the final product q. The production function is

$$q=g(x_3, u),$$

showing that the output of Division 1 becomes an input to Division 2. We denote the unit price of input x_3 by r_3 and the market price of the final product by p, and suppose that both are constant. Since Division 2 also operates as a profit center, its problem is:

$$\text{maximize } \pi_2=pq-r_3x_3-r_uu$$

$$\text{subject to the production function } q=g(x_3, u).$$

Finally we need to formulate the problem of the firm. Since Division 2 only sells in the market, the receipts of the firm are identical with the receipts of Division 2. To determine profit, we subtract total costs $C=r_1x_1+r_2x_2+r_3x_3$ from the receipts. The problem of the firm is therefore:

$$\text{maximize } \pi=pq-r_1x_1-r_2x_2-r_3x_3$$

$$\text{subject to the production functions } u=f(x_1, x_2) \text{ and } q=g(x_3, u).$$

Notice that the transfer price r_u does not appear in the problem of the firm because the cost of u to the firm is already taken care of by the term $r_1x_1+r_2x_2$.

Let us formulate the corresponding Lagrangian functions for the three problems:

$$L_1=r_uu-r_1x_1-r_2x_2-\lambda_1[f(x_1, x_2)-u],$$

$$L_2=pg(x_3, u)-r_3x_3-r_uu, \text{ and}$$

$$L =pg(x_3, u)-r_1x_1-r_2x_2-r_3x_3-\lambda[f(x_1, x_2)-u],$$

where λ_1 and λ represent Lagrangian multipliers. In order to find the optimum solutions to the three problems, we need to differentiate each Lagrangian function in terms of its variables and set each of the partial derivatives equal to zero:

Div. 1		*Firm*	
$\partial L_1/\partial u$	$=\partial(r_uu)/\partial u+\lambda_1=0$	$\partial L/\partial x_3$	$=pg_3-r_3=0$
$\partial L_1/\partial x_1$	$=-r_1-\lambda_1f_1=0$	$\partial L/\partial u$	$=pg_u+\lambda=0$
$\partial L_1/\partial x_2$	$=-r_2-\lambda_1f_2=0$	$\partial L/\partial x_1$	$=-r_1-\lambda f_1=0$
$\partial L_1/\partial \lambda_1$	$=f-u=0$	$\partial L/\partial x_2$	$=-r_2-\lambda f_2=0$
Div. 2		$\partial L/\partial \lambda$	$=f-u=0$
$\partial L_2/\partial x_3$	$=pg_3-r_3=0$		
$\partial L_2/\partial u$	$=pg_u-\partial(r_uu)/\partial u=0$		

where $f_i = \partial f/\partial x_i$, $g_u = \partial g/\partial u$, etc. As is well known, these conditions are only necessary for a maximum, but we shall not be concerned with second-order conditions here.

Let us see what needs to be done now. The firm's conditions consist of five equations and five variables (x_1, x_2, x_3, u, λ) and therefore, in principle, can be solved for the optimum values which will maximize the profit of the firm. We want to establish a correspondence between the equations of the firm and the equations of the divisions so that the optimum solutions to the divisional problems also solve the problem of the firm.

Before we do this, we need to know what the Lagrangian multipliers stand for. If $\partial L_1/\partial x_1$ is multiplied by dx_1, $\partial L_1/\partial x_2$ by dx_2, and both are added, we get

$$\begin{aligned} dC_1 = r_1 dx_1 + r_2 dx_2 &= -\lambda_1 (f_1 dx_1 + f_2 dx_2) \\ &= -\lambda_1 du \end{aligned}$$

and $-\lambda_1 = dC_1/du$. In words: The Lagrangian multiplier times (-1) represents the marginal cost of Div. 1. By performing the same operations with $\partial L/\partial x_1$ and $\partial L/\partial x_2$, we also get $-\lambda = dC_1/du$.

Now observe that the last three equations of the firm's first-order conditions agree with the last three equations of Div. 1. Moreover, $\partial L/\partial x_3$ is the same as $\partial L_2/\partial x_3$, and all we have to do is to establish a correspondence between $\partial L_1/\partial u$, $\partial L_2/\partial u$, and $\partial L/\partial u$. But this can easily be done by setting

$$\partial (r_u u)/\partial u = -\lambda_1 = -\lambda = MC_1 = dC_1/du$$

for then $\partial L_1/\partial u$ is identically satisfied, and Div. 2 will demand input u until the marginal revenue product pg_u equals the marginal cost of producing u, the same requirement imposed by the condition $\partial L/\partial u = 0$. In other words, with this choice the firm's optimal values x_1^0, x_2^0, x_3^0, u^0, λ^0 also satisfy the six divisional equations. Or, remembering the purpose of decentralization, when $\partial (r_u u)/\partial u = MC_1$, each division acting independently will operate at a level which maximizes the overall profit of the firm. For a specified value of u, Div. 1 determines the optimal values [4] of x_1, x_2, λ_1, while Div. 2 establishes the optimal x_3 and u. Thus, for given p and r_3, Div. 2 demands from Div. 1 the amount u^0 of the intermediate product, and Div. 1 in turn produces u^0 with the optimal input mix x_1^0, x_2^0.

How about the transfer price? If we integrate the last equation, we get

$$\begin{aligned} r_u u &= \int \frac{dC_1}{du}\, du + k \\ &= C_1(u) + k \end{aligned}$$

and the transfer price becomes

$$r_u = [C_1(u) + k]/u,$$

where k is an integration constant. Since $C_1(u)$ represents the total variable cost of Div. 1, we conclude that the transfer price equals the average variable cost of Div. 1, $C_1(u)/u$, plus (or minus) the term k/u. If k is set equal to total fixed cost, the transfer price will be the average cost. For most applications

[4] Since $\partial L_1/\partial u$ is an identity, Div. 1 has only three independent conditions.

average cost would seem to be the most appropriate pricing rule.[5] But since k is arbitrary, it can assume other values than total fixed cost if the situation requires it and still lead the divisions to the correct theoretical solution.[6] Thus, we could assign k a value of zero and transfer at average variable cost.[7] Because the constant k is arbitrary, we cannot speak of *the* transfer price, but merely of *a* transfer price. But in any event, the transfer price is some sort of an average. Why then did Hirshleifer determine MC_1 as the transfer price? Because he implicitly assumed that r_u does not depend on the output level of the intermediate product but is a constant. Note that when r_u is taken as a constant, we get $\partial(r_u u)/\partial u = r_u = MC_1$, the Hirshleifer solution. Can r_u be a constant? Yes, when, e.g., the production function $f(x_1, x_2)$ has the property of being homogeneous of degree 1. In this case the marginal cost of Div. 1 is a horizontal line and coincides with average variable cost. But this is only a special situation and the solution is not valid generally.[8] In particular, it is not valid for the Hirshleifer analysis because, when we examine his diagram once more, we see that he has drawn MC_1 as an increasing curve implying that marginal cost, and therefore r_u, depends on the output level.

Notice that when, e.g., r_u = average cost, we have no problems in the case of a declining marginal cost curve since average cost lies above marginal cost. Nor is there any necessity to restrict the output of a division as with the Hirshleifer solution. Division 2, being responsible for the total "profit" (except, perhaps, for a constant deduction k), has no incentive to search for a quasi-marginal revenue curve. The optimum output level for the firm is also the optimum for Division 2.

The result that the transfer price equals the average variable cost plus or minus k/u, is the general theoretical solution of the transfer pricing problem. It applies when there are more than two inputs and more than two divisions and, as we shall see in the next section, when there is a market for the intermediate product.

4. MARKET FOR THE INTERMEDIATE PRODUCT

We will assume that demand between the intermediate and the final product is independent, i.e., the volume of sales of one product does not affect the price of the other product. Let us first consider the case when the market prices of both products are constant. We designate the market price of the intermediate product by m, and use the equation $u = u_1 + u_2$ to indicate that of the total output u of the manufacturing division, u_1 is sold in the market and u_2 transferred to Division 2. The transfer price

---

[5] Ronen and McKinney also have suggested that average cost should be the appropriate transfer price. They adopt it to overcome the deficiencies of marginal cost pricing although they still hold that Hirshleifer's solution is the correct solution.

[6] One reader of this paper has suggested that the solution $r_u = [C_1(u) + k]/u$ is trivial since it is not *the* solution but allows many solutions by the arbitrariness of the constant k. That is a misunderstanding. What is relatively trivial is the mathematics used to derive the stated result. There is nothing very profound about an integration constant. But the fact that one appears does not render the result trivial. The constant simply tells us that we can adjust the transfer price to fit different circumstances and still have a correct solution.

[7] The Department of the Air Force several years ago decided to experiment with transfer pricing by requiring an Air Force test center to operate like an independent firm to the extent that this is possible in such an environment. The adjustment was a painful process as one would expect when an organization that had always been institutionally funded all of a sudden is ordered to act like a competitive firm. For a variety of reasons the experiment was not successful. One reason was that the facility was required to transfer close to average cost. Since a test facility has high fixed costs, it probably would have been more reasonable to initially require that transfers be priced at average variable cost until the managers and employees had become used to a radically different way of doing business.

[8] By casting the problem into a linear programming form, several authors have shown that marginal cost is the appropriate transfer price. But, of course, a linear programming formulation implies a linear homogenous production function.

is now designated by r_{u_2}. If otherwise the same relations and symbols are retained as before, we have the Lagrangian functions

$$L_1 = mu_1 + r_{u_2}u_2 - r_1x_1 - r_2x_2 - \lambda_1[f(x_1, x_2) - u] - \theta_1[u_1 + u_2 - u]$$

$$L_2 = pg(x_3, u_2) - r_3x_3 - r_{u_2}u_2$$

$$L = pg(x_3, u_2) + mu_1 - r_1x_1 - r_2x_2 - r_3x_3 - \lambda[f(x_1, x_2) - u] - \theta[u_1 + u_2 - u]$$

from which three sets of first-order conditions can be derived. Proceeding as in the last section, we obtain

$$r_{u_2} = m + \frac{k}{u_2} = \frac{[C_1(u_1 + u_2) + \phi(u_1)]}{u_2},$$

as the transfer price, where k and $\phi(u_1)$ are integration constants. The transfer price equals the market price plus or minus a term k/u_2. No purpose seems to be served in letting k be positive or negative. It would only introduce frictions between the divisions. The solution is therefore the same as Hirshleifer's: the transfer price is set equal to the market price of the intermediate product.

This solution does not change when the market price of the final product is no longer a constant. But when the price of the intermediate product also becomes a variable, the transfer pricing problem is more complicated. The transfer price is still the same but a division can no longer determine its optimum output without knowing the optimum output of the other division. This is obviously a problem of mutual dependence and the firm may prefer to have both divisions under single management.

5. TRANSFER PRICES IN THE PRESENCE OF CONSTRAINTS

Let us first consider a case where the head office for whatever reason wants the output of the manufacturing division not to be less than a certain quantity a^*, i.e., $u \geq a^*$. If we treat Div. 1 as a cost center, suppose that the intermediate product is not traded in the market, and take the final product price as a constant, we have the following objectives:

Div. 1 . . . minimize $C_1 = r_1x_1 + r_2x_2$
subject to $u = f(x_1, x_2)$

Div. 2 . . . maximize $\pi_2 = pq - r_3x_3 - r_uu$
subject to $q = g(x_3, u)$

Firm . . . maximize $\pi = pq - r_1x_1 - r_2x_2 - r_3x_3$
subject to $u = f(x_1, x_2)$, $q = g(x_3, u)$, and $u \geq a^*$.

We could, of course, impose the constraint $u \geq a^*$ directly on Div. 1, but unless we force Div. 2 in a similar way, 2 may buy a smaller quantity than a^* from 1. And the purpose of decentralization is to attain by the appropriate choice of transfer prices the objectives of the firm through suboptimization.

The corresponding Lagrangian functions for the above objectives are:

$$L_1 = r_1x_1 + r_2x_2 - \lambda_1[f(x_1, x_2) - u]$$

$$L_2 = pg(x_3, u) - r_3x_3 - r_uu$$

$$L = pg(x_3, u) - r_1x_1 - r_2x_2 - r_3x_3 - \lambda[f(x_1, x_2) - u] - \theta[u - a^* - s^2].$$

The constraint $u \geq a^*$ has been put in the form $u = a^* + s^2$ and incorporated into the function of the firm. The slack variable s^2 ensures that u cannot be less than a^*. As before, we derive the first-order conditions which, upon comparison, show that the transfer price is

$$r_u = \left[C_1(u) + \int \frac{d\pi}{da}\, du + k \right] / u.$$

To get a better understanding of the additional term $\int \frac{d\pi}{da}\, du$, consider Fig. 2, which shows the firm's profit as a function of the parameter a. It is assumed that the optimal u would be less than a^* if the constraint $u \geq a^*$ were eliminated. The level a^0 is obtained by setting $a^0 = u^0$, the optimal output without the constraint. When a^* does not exceed a^0, u would be set equal to the optimal a^0 and $\theta = d\pi/da = 0$ — the profit curve has zero slope over the interval $(0, a^0)$. When $a^* > a^0$, $\theta = d\pi/da < 0$ and $\int_{a^0}^{a^*} (d\pi/da)\, du$ represents the opportunity profit the firm gives up by requiring the manufacturing division to produce, and the distribution division to use, the amount $u = a^*$ instead of the optimal $u = a^0$. The integral is therefore negative and subtracted from the costs of Div. 1,

$$\int_{a^0}^{a^*} \frac{d\pi}{da}\, du = \pi(a^*) - \pi(a^0) < 0.$$

The opportunity profit function $\pi(a)$ can be easily computed from the first-order conditions of the firm. We conclude that the transfer price equals average cost less average opportunity profit,

$$r_u = \frac{C_1(u) + k}{u} - \frac{\pi(a^0) - \pi(a^*)}{u}.$$

Because of the lower transfer price to Div. 2 (lower relative to charging average cost), Div. 2 will demand a larger quantity of u (relative to a^0). More precisely, with this transfer price the amount a^*

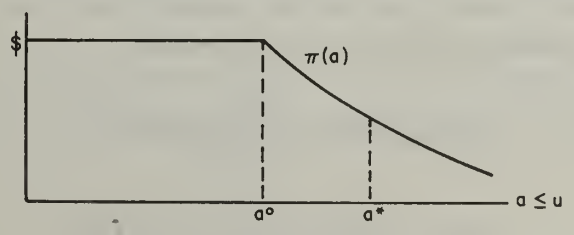


FIGURE 2

will be demanded by Div. 2 and without any direct constraints having been imposed on the divisions. Since Div. 1 is treated as a cost center, the head office may credit it with the opportunity profit $\pi(a^\circ) - \pi(a^*)$.

It follows that for the type of constraint we have been discussing, the transfer price is obtained by subtracting (or adding, as the case may be) the average opportunity profit from average cost.

Average cost can be adjusted in different ways, though, as we see by examining a case when the constraint is a budget constraint of the form $C = r_1x_1 + r_2x_2 + r_3x_3 \leq \bar{C}$, a constant. Let us suppose once again that the intermediate product is traded in the market. We shall assume that the firm's total costs would exceed the amount $\bar{C}$ for an optimal solution if there were no budget constraint. This eliminates the necessity of adding a slack variable to the inequality $C \leq \bar{C}$ as cost then equals the bound $\bar{C}$. From the appropriate Lagrangian functions, we get the following first-order conditions:

Div. 1	*Div.* 2	*Firm*
$\frac{\partial(mu_1)}{\partial u_1} - \theta_1 = 0$	$pg_3 - r_3 = 0$	$pg_3 - r_3(1+\eta) = 0$
$\frac{\partial(r_{u_2}u_2)}{\partial u_2} - \theta_1 = 0$	$pg_{u_2} - \frac{\partial(r_{u_2}u_2)}{\partial u_2} = 0$	$pg_{u_2} - \theta = 0$
		$\frac{\partial(mu_1)}{\partial u_1} - \theta = 0$
$-r_1 - \lambda_1 f_1 = 0$		$-r_1(1+\eta) - \lambda f_1 = 0$
$-r_2 - \lambda_1 f_2 = 0$		$-r_2(1+\eta) - \lambda f_2 = 0$
$f - u = 0$		$f - u = 0$
$u_1 + u_2 - u = 0$		$u_1 + u_2 - u = 0$
$\lambda_1 + \theta_1 = 0$		$\lambda + \theta = 0$
		$r_1x_1 + r_2x_2 + r_3x_3 - \bar{C} = 0$

Now $\eta = d\pi/d\bar{C}$, i.e., the Lagrangian multiplier η equals the change of profit per unit change in the budget. If the optimal C is larger than C without the constraint, η is positive: profit increases with an increasing budget. The magnitude of η can be determined from the first-order conditions of the firm.

Comparing the firm's conditions with the divisional conditions, we observe that if the divisions are required to use unit *standard* input costs of $(1+\eta)r_1$, $(1+\eta)r_2$, and $(1+\eta)r_3$ instead of r_1, r_2, and r_3, there is agreement between the global solution and the divisional solutions. Thus,

$$-\lambda_1 = \theta_1 = -\lambda = \theta = (1+\eta)MC_1,$$

where $MC_1 = (r_1 dx_1 + r_2 dx_2)/du$, and the transfer price becomes

$$r_{u_2}u_2 = (1+\eta)\int \frac{dC}{du}\, du_2 + \phi(u_1)$$

$$r_{u_2} = (1+\eta)[C_1(u_1+u_2) - C_1(u_1)]/u_2,$$

for $\phi(u_1) = -(1+\eta)C_1(u_1)$. Consequently, with a budget constraint the original average cost is multiplied by the opportunity profit rate $d\pi/d\bar{C}$.

As far as the divisions are concerned, costs are $(1+\eta)C_1(u)$ and $(1+\eta)C_2(q)$, and the intermediate product is transferred at an average cost. By adjusting the unit standard costs of the inputs in the indicated manner, the divisions will produce and transfer those quantities whose costs just equal the budget level. There is no need to impose direct constraints on the divisions. In fact it is likely that divisional budget constraints would be set at levels which are not in the best interests of the firm. Guidance through price adjustments should be the modus operandi of a decentralized corporation.

6. TRANSFER PRICING WITH AN ALTERNATIVE OBJECTIVE

Let us examine what happens when the firm's objective differs from profit maximization. Suppose that within a certain domain a firm is willing to trade off sales or revenues for profits, i.e., to increase revenues by a dollar the firm is willing to give up a certain amount of profit. Several such trade-off curves are shown in Figure 3, where a specific curve is defined for each level of utility. Consider, e.g., point A with revenue and profit of (R_A, π_A). Point B on the same curve has the same level of satisfaction to the firm as point A, implying that the increase in revenue $R_B - R_A$ is just worth the sacrifice in profit $\pi_A - \pi_B$.

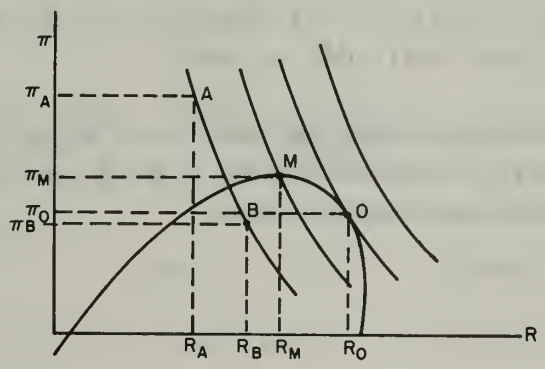


FIGURE 3

If we now sketch the profit-revenue function of the firm in the same diagram, we obtain a curve such as $F(R, \pi) = 0$.

Because the farther a trade-off curve is from the origin the higher the level of utility, the optimum occurs at point 0 where the curve F is just tangent to one of the indifference curves. The firm could maximize its profit at point M, but the trade-off curve through M has a lower utility than the trade-off curve through 0. Only when the firm does not exchange *any* increase in revenues for a dollar's worth of profit—the trade-off curves become horizontal straight lines—would the optimum occur at M.

Let us see what happens to the transfer price when a firm adopts the utility objective. We shall suppose that the divisions are designated as profit centers. If the intermediate product is traded in the market, the objectives of the divisions are the same as in section 4, while the firm's objective now becomes

$$\text{maximize } U = U(R, \pi)$$

subject to the constraints

$$\pi = pg(x_3, u_2) + mu_1 - r_1x_1 - r_2x_2 - r_3x_3$$

$$R = pg(x_3, u_2) + mu_1$$

$$u = f(x_1, x_2)$$

$$u = u_1 + u_2,$$

where $U(R, \pi)$ represents a utility function specifying the acceptable trade-offs between revenues and profits.

The corresponding Lagrangian functions are:

$$L_1 = mu_1 + r_{u_2}u_2 - r_1x_1 - r_2x_2 - \lambda_1[f(x_1, x_2) - u] - \theta_1[u_1 + u_2 - u]$$

$$L_2 = pg(x_3, u_2) - r_3x_3 - r_{u_2}u_2$$

$$L = U(R, \pi) - \lambda[f(x_1, x_2) - u] - \theta[u_1 + u_2 - u] - \eta[\pi - pg - mu_1 + r_1x_1 + r_2x_2 + r_3x_3] - \rho[R - pg - mu_1],$$

where λ, θ, η, ρ, etc., are Lagrangian multipliers, and $x_1, x_2, x_3, u_1, u_2, u, \pi, R$ variables. If, as usual, we assume that all functions can be differentiated, and we let the market price of the final commodity be a constant, the following first-order conditions hold:

Div. 1	*Div. 2*	*Firm*
$\frac{\partial(mu_1)}{\partial u_1} - \theta_1 = 0$	$pg_3 - r_3 = 0$	$U_R - \rho = 0$
$\frac{\partial(r_{u_2}u_2)}{\partial u_2} - \theta_1 = 0$	$pg_{u_2} - \frac{\partial(r_{u_2}u_2)}{\partial u_2} = 0$	$U_\pi - \eta = 0$
$-r_1 - \lambda_1 f_1 = 0$		$(\eta + \rho)pg_3 - \eta r_3 = 0$
$-r_2 - \lambda_1 f_2 = 0$		$(\eta + \rho)pg_{u_2} - \theta = 0$
$f - u = 0$		$(\eta + \rho)\frac{\partial(mu_1)}{\partial u_1} - \theta = 0$
$u_1 + u_2 - u = 0$		$-\eta r_1 - \lambda f_1 = 0$
$\lambda_1 + \theta_1 = 0$		$-\eta r_2 - \lambda f_2 = 0$
		$f - u = 0$
		$u_1 + u_2 - u = 0$
		$\lambda + \theta = 0$
		$\pi - pg - mu_1 + r_1x_1 + r_2x_2 + r_3x_3 = 0$
		$R - pg - mu_1 = 0.$

Comparing the divisional equations with the firm's equations, we notice the following desired and actual correspondences:

$$\theta_1 = \theta/(\eta+\rho)$$

$$\lambda_1 = \lambda/(\eta+\rho)$$

$$-\lambda_1 = -\lambda/\eta = MC_1 = dC_1/du.$$

Therefore,

$$\frac{\partial(r_{u_2}u_2)}{\partial u_2} = \frac{\theta}{\eta+\rho} = \frac{-\lambda}{\eta+\rho}$$

(from the first-order condition $\lambda+\theta=0$ of the firm)

$$= \frac{\eta MC_1}{\eta+\rho} \qquad \text{(from the last of the noted correspondences)}$$

and

$$r_{u_2}u_2 = \frac{\eta}{\eta+\rho}\int \frac{dC_1}{du}\,du_2 + \phi(u_1) = \left(\frac{U\pi}{U_\pi+U_R}\right) C_1(u_1+u_2) + \phi(u_1)$$

determines a transfer price of

$$r_{u_2} = \left(\frac{U_\pi}{U_\pi+U_R}\right)[C_1(u_1+u_2) - C_1(u_1)]/u_2.$$

Observe that when each of the input prices r_1, r_2, and r_3 is multiplied by the factor $\eta/(\eta+\rho)$ so that the standard unit input costs to the divisions become $\eta r_i/(\eta+\rho)$ $(i=1, 2, 3)$, each division acting as a profit center will generate the solution desired as optimal by the firm. As far as the divisions are concerned, the cost of the intermediate product is $\eta C_1/(\eta+\rho)$ after adjustment and they are transferring once more at an average cost. Although the divisions are not subjected to the direct constraints $r - pg - mu_1 + r_1x_1 + r_2x_2 + r_3x_3 = 0$, $R - pg - mu_1 = 0$, these equations will be automatically satisfied with the η, ρ specified at their optimal values (which are computed from the first-order conditions of the firm).

Since in our diagram $U_\pi, U_R \geq 0$, it follows that $U_\pi/(U_\pi+U_R) = \eta/(\eta+\rho) \leq 1$. Thus the effect of the adjustment is to reduce the original costs, which, of course, implies that the divisions will sell and use more of the intermediate product. Notice that when the firm is a profit maximizer and derives no satisfaction from increased sales if this means a decrease in profits, $U_R=0$, $U_\pi/(U_\pi+U_R)=1$, and we get the profit maximizing solution.

7. CONCLUSIONS

We have shown that an average cost is the correct theoretical transfer price under a variety of conditions. When constraints are included which are normally always present in any real case, the transfer price becomes an adjusted average cost. It is obviously difficult and perhaps often impossible to adjust prices precisely in practice. But if there is a commitment to the principles of transfer pricing, the approximate magnitudes of the required adjustments can be determined to give Adam Smith's invisible hand a fair chance to guide the divisions toward the optimum of the firm.

BIBLIOGRAPHY

[1] Arrow, K. J., "Optimization, Decentralization, and Internal Pricing in Business Firms," Western Data Process Center, University of California, Los Angeles, pp. 9–18 (1959).

[2] Berkwitt, G. J., "Do Profit Centers Really Work?" Dun's Review *93*, 29–31, 117 (1969).

[3] Bierman, H., Jr., "Pricing Intracompany Transfers," Accounting Review *34*, 429–432 (1959).

[4] Cook, P. W., Jr., "Decentralization and the Transfer Price Problem," The Journal of Business *28*, 87–94 (1955).

[5] Cook, P. W., Jr., "New Technique for Intracompany Pricing," Harvard Business Review *35*, 74–80 (1957).

[6] Dean, J., "Decentralization and Intercompany Pricing," Harvard Business Review *33*, 65–74 (1955).

[7] Dopuch, N. and D. F. Drake, "Accounting Implications of a Mathematical Programming Approach to the Transfer Price Problem," Journal of Accounting Research *2*, 10–24 (1964).

[8] Gould, J. R., "Internal Pricing in Firms When There Are Costs Of Using an Outside Market," The Journal of Business *37*, 61–67 (1964).

[9] Greer, H. C., "Divisional Profit Calculation—Notes on the 'Transfer Price' Problem," N.A.A. Bulletin *43*, 5–12 (1962).

[10] Hass, J. E., "Transfer Pricing in a Decentralized Firm," Management Science *14*, B-310–331 (1968).

[11] Hirshleifer, J., "On the Economics of Transfer Pricing," The Journal of Business *29*, 172–184 (1956).

[12] Hirshleifer, J., "Economics of the Divisionalized Firm," The Journal of Business *30*, 96–108 (1957).

[13] Hirshleifer, J., "Internal Pricing and Decentralized Decisions," in *Management Controls—New Directions in Basic Research*, ed. by C. P. Bonini, R. K. Jaedecke, H. M. Wagner (McGraw-Hill, N.Y., 1963).

[14] Menge, J. A., "The Backward Art of Interdivisional Transfer Pricing," Journal of Industrial Economics *9*, 215–232 (1961).

[15] Naert, P. A. and T. L. Janssen, "On Suboptimization in Decomposition Approaches to Transfer Pricing," Journal of Industrial Economics *19*, 220–230 (1971).

[16] Onsi, M., "A Transfer Pricing System Based on Opportunity Cost," The Accounting Review *45*, 535–543 (1970).

[17] Ronen, J. and G. McKinney, III, "Transfer Pricing for Divisional Autonomy," Journal of Accounting Research *8*, 99–112 (1970).

[18] Samuels, J. M., "Opportunity Costing: An Application of Mathematical Programming," Journal of Accounting Research *3*, 182–191 (1965).

[19] Shillinglaw, G., *Cost Accounting: Analysis and Control*, Rev. Ed. (Irwin Homewood, III, 1967), chaps. 25, 26.

[20] Solomons, D., *Divisional Performance: Measurement and Control* (Irwin Homewood, III, 1965).

ISOQUANTS OF CONTINUOUS PRODUCTION CORRESPONDENCES

G. Bol and O. Moeschlin

University of Karlsruhe
Federal Republic of Germany

ABSTRACT

Bol has discussed consequences of the continuity of production correspondences in connection with relations between efficient input and output vectors. Isoquants of continuous production correspondences are used here to extend this work. Simplifications to existing theory are discussed.

INTRODUCTION

In an earlier paper [2] Bol discussed consequences of the continuity of production correspondences in connection with relations between efficient input and output vectors. Thereby it becomes clear, that the intuitively expected results can only be established in the case of one input and/or for one output. Stimulated by Shephard, in the present paper we deal with a similar problem as in [2] substituting the sets of efficient input and output vectors by the isoquants of production correspondences. Directly connected with this problem is the existence of a joint production function [6]. Following this line of approach, we are led to a counter example of an existence theorem for joint production functions given by Shephard. Although only a weak additional assumption is to be made, a corrected and improved version of the proof brings a considerable simplification of the formal treatment; moreover some of the assumptions originally made by Shephard can be dropped. As it will be shown, the converse of the theorem is also true. The existence of a joint production function to a given production correspondence entails the continuity of the latter.

1. CONTINUOUS PRODUCTION CORRESPONDENCES[1]

The technical possibilities of a firm, which transforms n input commodities in m output commodities, may be described by a mapping $P\colon R_+^n \to 2^{R_+^m}$[2] which assigns to each input vector $x \in R_+^n$ the set $P(x)$ of output vectors attainable by x. From P we get a mapping $L\colon R_+^m \to 2^{R_+^n}$, which assigns to any output vector u the set $L(u)$ of input vectors, by which u can be produced. P and L are inverse correspondences, i.e., $u \in P(x)$ if and only if $x \in L(u)$. For P and L we assume some "natural" properties; in order to keep our approach transparent, we tried to confine to just those assumptions, which really are needed.[3]

---

[1] Continuity of production correspondences are explicitly discussed in [1] and [3]. Especially in [3] connections to the different topologies on systems of subsets of a topological space are shown. Here we give only some results without proofs.

[2] For a set M 2^M denotes the set of subsets of M.

[3] They are essentially the same as in [5].

1.1. Definition

Let $P: R_+^N \to 2^{R^m_+}$ and $L: R_+^m \to 2^{R^n_+}$ be inverse correspondences ($u \in P(x) \iff x \in L(u)$). The pair (P, L) is called a *production system*, if P and L fulfill the following properties:

1. For all $x \in R_+^n$ $P(x) \neq \phi$, for all $u \in R_+^m$ $L(u) \neq \phi$;
2. For all $x \in R_+^n$ $P(x)$ is compact, for all $u \in R_+^m$ $L(u)$ is closed;
3. For all $x, x' \in R_+^n$ holds $P(x) \subset P(x')$ if $x \leqq x'$[4];
4. For all $u, u' \in R_+^m$ holds $L(u) \supset L(u')$ if $u \leqq u'$;
5. For all $u_0 \in R_+^m$ there exists a neighborhood U of u_0 in R_+^m and a compact set $K \subset R_+^m$ with

$$E_L(u) := \{x \in L(u) \mid : x' \in L(u) : x' \nexists x' \in L(u) : x' \leqq x\} \subset K \qquad \text{for all } u \in U.$$

P is also called a production correspondence and L its inverse correspondence.

An explicit interpretation is given in the literature (especially in [5], [6], and [7]). Axiom 5 is a stronger form of an assumption, which Shephard calls "assymetric property" and explains it as a restriction on substitutability [7]: "No output is attained efficiently by unbounded application of any input, as substitute for another perhaps."

Here Axiom 5 means that the same is also valid, if "small" variations of the output vector are admitted.

1.2. Lemma

For a production system (P, L), the graph of P (which is the same as the graph of L): $\{(u, x) \mid u \in P(x), x \in R_+^n\} = \{(u, x) \mid x \in L(u), u \in R_+^m\}$ is closed.[5]

1.3. Theorem

For a production system (P, L), P is continuous in $x_0 \geqq 0$ with respect to the Hausdorff-topology)[6] on the set of closed subsets of R_+^m if and only if

$$(*)\ P(x_0) = \overline{\bigcup_{0 \leqq x \hat{<} x_0} P(x)}.$$

Moreover P is continuous in 0.

---

[4] For two vectors $x, x' \in R^k$ we use the following notations:

$$x < x' \iff x_i < x'_i \quad \text{for } i = 1, \ldots, k$$

$$x \hat{<} x' \iff x_i < x'_i \quad \text{or} \quad x_i = x'_i = 0 \quad \text{for } i = 1, \ldots, k$$

$$x \leqq x' \iff x_i \leqq x'_i \quad \text{for } i = 1, \ldots, k$$

$$x \leqslant x' \iff x \leqq x' \quad \text{and} \quad x \neq x'$$

[5] In Shephard's notation this is equivalent to upper-semicontinuity of P and L.

[6] We omit here the definition of the Hausdorff-topology, given for example in [3]. In the following, we use continuity of P or L only in the form of the properties (*) or (**) and corollary 1.5, so the reader need not be familiar with this topology.

1.4. Theorem

For a production system (P, L) L is continuous in u_0 with respect to the Hausdorff-topology if and only if

$$(**)\ L(u_0) = \overline{\bigcup_{u > u_0} L(u)}.$$

1.5. Corollary

For a production system (P, L) we have:

1. If P is continuous in x_0 and (x_n) a sequence in R^n_+ with $\lim x_n = x_0$, then there exists for all $u_0 \in P(x_0)$ a sequence (u_n) in R^m_+ with $u_n \in P(x_n)$ and $\lim u_n = u_0$;

2. If L is continuous in u_0 and (u_n) a sequence in R^m_+ with $\lim u_n = u_0$, then there is for all $x_0 \in L(u_0)$ a sequence (x_n), $x_n \in L(u_n)$, with $\lim x_n = x_0$.

1.6. Definition

A production system (P, L) is called *continuous*, if P and L are continuous in all $x \in R^n_+$ resp. $u \in R^m_+$ with respect to the Hausdorff-topology.

2. ISOQUANTS

2.1. Definition [6]

For a production system (P, L) and an input vector $x \in R^n_+$ with $P(x) \neq \{0\}$ the set

$$I_P(x) = \{u \in P(x) \mid \exists\, \lambda > 1 : \lambda \cdot u \in P(x)\}$$

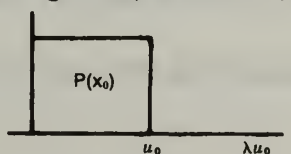
is called the *output isoquant* of x. For $x \in R^n_+$ with $P(x) = \{0\}$ we set $I_P(x) = \{0\}$. Analogously, for an output vector $u \geqq 0$ the set $I_L(u) = \{x \in L(u) \mid \exists\, 0 \leqq \mu < 1 : \mu \cdot x \in L(u)\}$ is called the *input isoquant* of u. For $u = 0$ we set $I_L(0) = \{0\}$.

2.2. Lemma

For a production system (P, L) the input isoquant $I_L(u)$ is closed for all u.

PROOF: Let $u \neq 0 - I_L(0) = \{0\}$ is closed — and let (x_n) be a sequence in $I_L(u)$ with $\lim x_n = x_0$. Assuming $x_0 \notin I_L(u)$ there exists a λ with $0 \leqq \lambda < 1$ and $\lambda x_0 \in L(u)$. Because $\lambda x_0 \hat{<} x_0$ and $\{x \mid \lambda x_0 \leqq x\} \subset L(u)$, there exists a neighborhood U of x_0 with $U \cap R^n_+ \subset L(u)$.[7] But $\lim x_n = x_0$, and therefore $x_n \in U$ for almost all n. So x_n is an inner point of $L(u)$ relative to R^n_+, in contradiction to $x_n \in I_L(u)$.

REMARK: Output isoquants are, in general, not closed, as the following example illustrates:



[7] For example $U = \{x \mid \|x - x_0\|_\infty < \epsilon\}$, where $\epsilon = \min\{(1-\lambda)x_{0,i} \mid i = 1, \ldots, n \text{ and } x_{0,i} > 0\}$.

This example works because for $\lambda > 1$, $\lambda u_0 \in P(x_0)$ does not imply the existence of a neighborhood of u_0 in $P(x_0)$ relative to R_+^m.

2.3. Definition

A production system with $P(x)$ convex for all x is called *production system with convex output structure.* (c.o.s. – production system).

2.4. Lemma

For a c.o.s. – production system (P, L) $I_P(x)$ is closed for all $x \in R_+^n$.

PROOF: First we show that convexity of $P(x)$ implies a weaker property (*), which is sufficient for the proof of Lemma 2.4.

1. If $P(x)$ is convex, then: (*) Denoting $\{i \in \{1, \ldots, m\} \mid \exists\ u \in P(x) : u_i > 0\}$ by I, for all $\bar{u} \in P(x)$ and for all $\epsilon > 0$ there exists a $u' \in P(x)$ with $\| u' - \bar{u} \|_\infty < \epsilon$ and $u_i' > 0$ for all $i \in I$.

Because $P(x)$ is convex, there exists an $u^0 \in P(x)$ with $u_i^0 > 0$ for all $i \in I$. For $\bar{u} \in P(x)$ we get $u^\lambda = \lambda u^0 + (1 - \lambda)\bar{u} \in P(x)$ for all λ with $0 < \lambda \leqq 1$ and $u_i^\lambda > 0$ for all $i \in I$. Then for small λ,

$$\| u^\lambda - \bar{u} \|_\infty < \epsilon.$$

2. Let $u^n \in I_P(x) \neq \{0\}$ be a sequence with $\lim u^n = \bar{u}$. If $\bar{u} \notin I_P(x)$, there is a $\lambda > 1$ with $\lambda \bar{u} \in P(x)$.

Because of property (*), for all $\epsilon > 0$ there is a $u_\epsilon' \in P(x)$ with $\| u_\epsilon' - \lambda \bar{u} \|_\infty < \epsilon$ and $u_{\epsilon,i}' > 0$ for all $i \in I$. Because $P(x) \subseteq \{u \notin R_+^m \mid u_i = 0$ for all $i \notin I\}$, we can assume without loss of generality that $I = \{1, \ldots, m\}$ and therefore $\bar{u} < u'$. Then since $\lim u^n = \bar{u}$, $u^n < u'$ for almost all n, in contradiction to $u^n \in I_P(x)$.

2.5. Lemma

For a continuous production system (P, L) I_p and I_L are lower-semicontinuous, i.e. for (x_k), $x_k \in R_+^n$, with $\lim x_k = x_0$ and $u_0 \in I_p(x_0)$ there exists for almost all k a $u_k \in I_p(x_k)$ with $\lim u_k = u_0$ (analogously for I_L).

PROOF:

1. For a sequence (x_k) in R_+^n with $\lim x_k = x_0$ and $0 \neq u_0 \in I_p(x_0)$ there is – P is continuous – a sequence $u_k \in P(x_k)$ with $\lim u_k = u_0$.

Let $\lambda_k = \max \{\lambda \mid \lambda u_k \in P(x_k)\}$ if $P(x_k) \neq \{0\}$ and $\lambda^k = 1$ otherwise. Then $\lambda_k \geqq 1$ and the sequence (λ_k) is bounded, because $P(C) = \bigcup_{x \in C} P(x)$ is compact for compact $C \subset R^n$. Therefore (λ_k) converges to 1, if every converging subsequence converges to 1. We assume λ_{ki} converges and $\lim \lambda_{ki} = \lambda_0$. Then $\lambda_0 \geqq 1$ and $\lim \lambda_{ki} u_{ki} = \lambda_0 u_0 \in P(x_0)$ by Lemma 1.2. and $\lambda_0 \leqq 1$ by $u_0 \in I_P(x_0)$.

For $u_0 = 0$ we have $P(x_0) = \{0\}$. For arbitrary $u_k \in P(x_k)$ and any converging subsequence u_{k_i}, u_{k_i} converges to 0 by Lemma 1.2. Because (u_k) is bounded, (u_k) converges to 0 and we can choose $u_k \in I_P(x_k)$ arbitrarily.

2. The proof for the lower-semicontinuity of L is analogous.

2.6. Remark

In the next paragraph we give a counter example for a theorem of Shephard. This example shows

also that the continuity of the correspondence I_L does not follow from the continuity of the production system.

Another example, a quite simple one, is the production correspondence $P: R_+^2 \to 2^{R_+^2}$ with $P(x_1, x_2) = \{(u_1, u_2) \mid 0 \leqq (u_1, u_2) \leqq (x_1, x_2)\}$ (then $L(u_1, u_2) = \{(x_1, x_2) \mid (u_1, u_2) \leqq (x_1, x_2)\}$). For $(x_1^n, x_2^n) = \left(1 - \frac{1}{n}, \frac{1}{n}\right)$ the vector $\left(\frac{1}{2}, \frac{1}{n}\right) \in I_P(x_1^n, x_2^n)$ but $\left(\frac{1}{2}, 0\right) \in I_P(1, 0)$. Similarly, I_L is not continuous in points (u_1, u_2) with $u_1 = 0$ or $u_2 = 0$.

3. EXISTENCE OF A JOINT PRODUCTION FUNCTION

3.1. Definition (~6≦)

For a production system (P, L) a function $F : R_+^m \times R_+^n \to R$ is called *joint production function*, if

(1) $$\text{for } u \geq 0, \qquad x \in I_L(u) \Leftrightarrow F(u, x) = 0$$

(2) $$\text{for } x \geq 0, \qquad u \in I_P(x) \Leftrightarrow F(u, x) = 0.$$

3.2. Lemma

For a production system (P, L) *there exists a joint production function if and only if*

(*) for all $x \geq 0$ with $P(x) \neq \{0\}$ and $u \geq 0$ we have $u \in I_P(x) \Leftrightarrow x \in I_L(u)$.

PROOF: Let $F : R_+^m \times R_+^n \to R$ be a joint production function, $x \geq 0$ with $P(x) \neq \{0\}$, then $u \in I_P(x) \Leftrightarrow F(u, x) = 0 \Leftrightarrow x \in I_L(u)$. If (*) holds,

$$\{(u, x) \mid x \geq 0, \qquad u \geq 0 : P(x) \neq \{0\}, \qquad u \in I_p(x)\} =$$

$$\{(u, x) \mid x \geq 0, \qquad u \geq 0 : P(x) \neq \{0\}, \qquad x \in I_L(u)\} =: M$$

and we set

$$F(u, x) = \begin{cases} 0 & (u, x) \in M \\ 0 & u = 0 \text{ and } P(x) = \{0\} \\ 1 & \text{otherwise.} \end{cases}$$

Routine checking shows, that F is a joint production function for (P, L).

3.3. Theorem

If a joint production function exists for a production system (P, L), then (P, L) is continuous.

PROOF:

a. Let us assume that P is discontinuous in $x_0 \geq 0$, then $P(x_0) \neq \bigcup_{0 \leq x \hat{<} x_0} P(x) =: \tilde{P}(x_0)$ and there exists a $u_0 \in P(x_0)$ with $u_0 \notin \tilde{P}(x_0)$ (and therefore $u_0 \neq 0$). Without loss of generality we assume $u_0 \in I_P(x_0)$. Because $\tilde{P}(x_0)$ is closed, there is a neighborhood U of u_0 with $U \cap \tilde{P}(x_0) = \emptyset$. Thus there is a λ with $0 < \lambda < 1$ and $\lambda u_0 \notin \tilde{P}(x_0)$ (but $\lambda u_0 \in P(x_0)$ by $\lambda u_0 \leqq u_0$). For all $x_0 \geq 0$ and all μ with $0 < \mu < 1$ we have $0 \leqq \mu x_0 \hat{<} x_0$ and it follows that $\lambda u_0 \notin P(\mu x_0)$ and hence $x_0 \in I_L(\lambda u_0)$. On the other hand, $\lambda u_0 \in I_P(x_0)$ and from the Lemma above it follows that there is no joint production function.

b. Assuming L is a discontinuous in a point u_0, an analogous argument leads to a contradiction.

REMARK: As the proof shows, the following part of assumption (*) in 3.2. is sufficient for the continuity of P (resp. L): $\forall x \geq 0$: $P(x) \neq \{0\}$ and $\forall u \geq 0$: $x \in I_L(u) \Rightarrow u \in I_P(x)$ (resp. $u \in I_P(x) \Rightarrow x \in I_P(u)$).

The converse of the theorem is not true in general. Under stronger assumptions (assumptions about attainability and convexity) Shephard gives a proof in [6] for the existence of a joint production function for continuous production systems. But his proof is wrong in one point (for $x \geq 0$ and $0 < \lambda < 1$ we do not necessarily have $\lambda x < x$). The example we give below fulfills all assumptions made by Shephard, so it is also a counter example for his theorem.

3.4. Counter Example for the Converse of Theorem 3.3. (and the Corresponding Theorem of Shephard).

For $u \in R_+$ (one output-commodity) $L(u) \subset R^2_+$ (two input-commodities) is given by:

$$L(u) = \begin{cases} \{(x_1, x_2) \in R^2_+ \mid u \leqq x_1, \quad x_1 + x_2 \geqq 2u\} & 0 \leqq u \leqq 1 \\ \{(x_1, x_2) \in R^2_+ \mid u \leqq x_1, \quad x_1 + x_2 \geqq 2u, \quad ux_2 + (u-1)x_1 \geqq 2u^2 - 2\} & 1 \leqq u \leqq 2 \\ \{(x_1, x_2) \in R^2_+ \mid u \leqq x_1, \quad 2x_2 + x_1 \geqq 3u\} & 2 \leqq u. \end{cases}$$

Figure 1 illustrates the isoquants $I_L(u)$ for some special values of u:

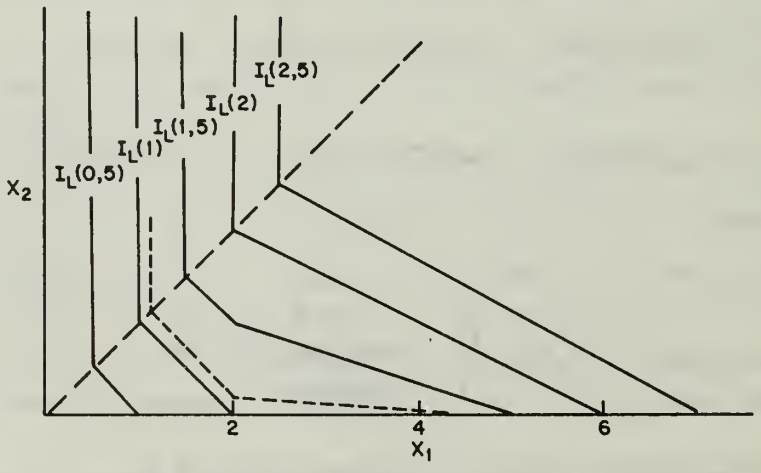


FIGURE 1

The production correspondence P to L can be calculated from L as $P(x) = [0, f(x)]$ with $(I_P(x) = \{f(x)\})$:

$$f(x) = \begin{cases} x_1 & x_2 \geqq x_1 \quad \text{(I)} \\ 1/2(x_1 + x_2) = \{f(x)\}: & x_1 > x_2,\ x_1 \leqq 2 \quad \text{(II)} \\ 1/4[(x_1 + x_2) + \sqrt{(x_1 + x_2)^2 - 8x_1 + 16}] & x_1 > x_2,\ 6 \geqq x_1 \geqq 2,\ x_2 + 1/2x_1 \leqq 3 \quad \text{(III)} \\ 1/3(x_1 + 2x_2) & x_1 > x_2,\ x_1 \geqq 2,\ x_2 + 1/2x_1 \geqq 3 \quad \text{(IV)} \end{cases}$$

(P, L) fulfills the properties 1 to 5. P is continuous, if f is continuous. But f is obviously continuous in the domains I to IV of Figure 2 and the different definitions yield the same value at the common boundary of two domains. So f is continuous.

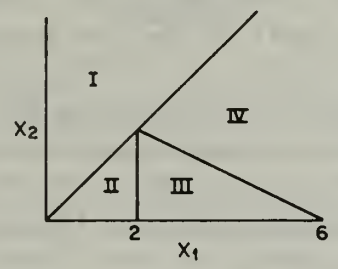


FIGURE 2

To show that L is continuous is more technically complicated, but straightforward and can be omitted here.

So (P, L) is a continuous production correspondence, but there is no joint production function because of $1 \in I_P(4, 0)$ but $(4, 0) \notin I_L(1)$ $((2, 0) \in L(1))$.

REMARK: The example shows, that the crucial point is the possible existence of an $x \not> 0$ with $P(x) \neq \{0\}$. If we exclude this, we can give a simple proof, which in contrast to Shephard's proof does not require any separation theorem (and therefore no convexity assumptions) nor any assumption about attainability.

3.5. Definition

The input commodities of a production system (P, L) are called *weakly coupled*,[8] if $P(x) \neq \{o\}$ only for $x > 0$.

3.6. Theorem

For a continuous production system (P, L) with weakly coupled input commodities a joint production function exists.

PROOF:

1. $\forall x \geqq 0$ with $P(x) \neq \{0\}, \forall u \geqq 0$: $x \in I_L(u) \Rightarrow u \in I_P(x)$. We assume that there is an $x_0 \geqq 0$ with $P(x) \neq \{0\}$ and an $u_0 \geqq 0$ such that $x_0 \in I_L(u_0)$ but $u_0 \notin I_P(x_0)$ and hence a $\lambda_0 > 1$ with $\lambda_0 u_0 \in P(x_0)$. By

[8] Compare *W. Krelle* [4], p. 12. Sometimes the expression is used: "All inputs are individually essential". (In [6] for the one-commodity-production).

continuity of P, we have, for a sequence (μ_n) of real numbers with $0 \leq \mu_n < 1$ and $\lim \mu_n = 1$, a sequence (u_n), $u_n \in P(\mu_n x_o)$, with $\lim u_n = \lambda_0 u_0$. But then, for almost all n, $u_n \hat{>} u_0$ because of $\lambda_0 u_0 \hat{>} u_0$. (Compare footnote 7). Therefore we have $u_0 \in P(\mu_n x_0)$ for almost all n, in contradiction to $x_0 \in I_L(u_0)$ since $\mu_n < 1$.

2. $\forall x \geq 0$ with $P(x) \neq \{0\}$ and $\forall u \geq 0$: $u \in I_P(x) \Rightarrow x \in I_L(u)$. As the input commodities are weakly coupled, it follows from $P(x_0) \neq \{0\}$ that $x_0 > 0$. For a sequence (x_n) with $\lim x_n = \lambda x_0$, $0 < \lambda < 1$, $\lambda x_0 < x_0$ [9] implies that for almost all n $x_n < x_0$. The rest of the proof is essnetially the same as in 1.

REFERENCES

[1] Bol, G., "Stetigkeit und Effizienz bei mengenwertigen Produktionsfunktionen," Mathematical Systems in Economics Vol. 7, Anton Hain Verlag (1973).

[2] Bol, G., "Continuity of Production Correspondences and a Relation Between Efficient Input and Output Vectors," in W. Eichhorn and R. W. Shephard, Seminar in Production Theory, Karlsruhe (1973).

[3] Bol, G., "Topologien auf Systemen von Teilmengen des R^n und Auswirkungen auf Produktionskorrespondenzen," Zeitschrift für Nationalökonomie, *34*, 91–106 (1974).

[4] Krelle, W., "Produktionstheorie," Tübingen (1969).

[5] Opitz, O., Zum technischen Optimierungsproblem des Unternehmers," Schweizerische Zeitschrift für Volkswirtschaft und Statistik (1970).

[6] Shephard, R. W., *Theory of Cost and Production Function* (Princeton University Press, 1970).

[7] Shephard, R. W., "Indirect Production Functions," Mathematical Systems in Economics, Vol. 10, Anton Hain Verlag, (1974).

---

[9] $\lambda x_0 \hat{<} x_0$ is not enough.

AN EFFICIENT CYCLIC COORDINATE METHOD FOR OPTIMIZING PENALTY FUNCTIONS

Mokhtar S. Bazaraa

School of Industrial and Systems Engineering
Georgia Institute of Technology

ABSTRACT

In this study, a simple and efficient cyclic coordinate search procedure is used to optimize penalty functions. Since the contours of the penalty function are very ill-behaved, an accurate line search is very difficult to achieve. Due to accumulated errors in line search, this makes "simple-minded" search directions just as good as more sophisticated directions, and actually better since they require a smaller effort per iteration. Of course this is only true if the search procedure is able to "ride" along steep ridges, and meanwhile move an appreciable distance towards the optimal, if at all possible. Computational results on the cyclic coordinate method seems to support this point of view, and shows robustness, reliability, and efficiency of the method.

1. INTRODUCTION

Consider the following nonlinear programming problem P:

$$\text{Minimize } f(x)$$

Subject to:

$$g_i(x) \leq 0 \qquad i=1, 2, \ldots, m$$

$$h_i(x) = 0 \qquad i=1, 2, \ldots, k.$$

One of the popular schemes for solving the above problem is a penalty or barrier function approach, where the constrained problem is transformed into a single unconstrained problem, or a sequence of unconstrained problems (see [2, 5, 8]). At least for the case of parametric penalty and barrier functions, however, the resulting unconstrained problem is very ill-conditioned (see for example [5]) such that, even the most efficient unconstrained minimizing techniques have a great deal of difficulty in handling the unconstrained penalty problem(s). It is therefore of paramount importance to devise unconstrained techniques which are especially designed or modified to handle penalty and barrier functions (see [4]).

In this study, the difficulties involved with optimizing parametric penalty functions are reexamined. It turns out that some of the most elementary procedures for unconstrained optimization, which are usually regarded in the literature as inefficient, can be successfully adopted to solve the penalty problem. In fact, a "cyclic coordinate method" has been coded and proved efficient for solving constrained optimization problems, by the aid of penalty functions. Nine standard test problems with varying de-

grees of difficulty have been attempted. The algorithm was able to solve all the test problems. The computational time is comparable to, and in many cases is significantly less than, the computational time used by some of the successful penalty and nonpenalty based nonlinear programming codes.

2. SOME REMARKS ON OPTIMIZING PENALTY FUNCTIONS

Solving the constrained nonlinear programming problem P via a parametric penalty function technique, requires finding:

$$\sup_{\lambda} \operatorname*{Inf}_{x \epsilon R_n} (f(x) + \lambda I(x)) = \lim_{\lambda \to \infty} \operatorname*{Inf}_{x \epsilon R_n} (f(x) + \lambda I(x)),$$

where I is a function which measures the infeasibility of x, i.e., $I(x) = 0$ if $g_i(x) \leq 0$ for $i = 1, 2, \ldots, m$ and $h_i(x) = 0$ for $i = 1, 2, \ldots, k$, and $I(x)$ is positive otherwise. A typical function I is given below:

$$I(x) = \sum_{i=1}^{m} (\max\ (0, g_i(x)))^2 + \sum_{i=1}^{k} h_i^2(x)$$

see for example [1, 8].

In order to solve the above problem, one must solve an unconstrained problem of the form

$$\operatorname*{Minimize}_{x \epsilon R_n} f(x) + \lambda I(x)$$

where γ is a large positive number. Whether this is done in a single step, or through a sequence of problems of increasing λ's, one must overcome the difficulties associated with a large penalty parameter. In fact, if during the search process, a feasible (or near feasible) point x is at hand, it becomes very difficult to make any movement away from this point, which usually results in premature termination. Given any direction d with norm 1, and since the ridge consisting of feasible points $\{x : I(x) = 0\}$ is usually highly nonlinear, then $I(x + \mu d) > 0$ for any μ. Since $I(x + \mu d)$ is weighted by a very large λ, then

$$f(x + \mu d) + \lambda I(x + \mu d) > f(x) + \lambda I(x) = f(x),$$

for any μ of "reasonable length." This is due to the fact that any improvement in the value of f will be offset by the increase in the infeasibility, which is weighted by λ. Improving on x is thus only realizable by moving through very small values of μ, which leads to either premature termination, or very slow progress. This discussion explains the difficulty with optimizing penalty functions, namely, the contours of the penalty function for a larger value of λ, are very ill-conditioned, and have a steep "dip" along the ridge $\{x: I(x) = 0\}$, which makes it very difficult to find the minimum over this ridge.

This point is further illustrated by Rosenbrock's function (see [6]), $(1 - x_1)^2 + 100(x_2 - x_1^2)^2$. This function has been used in the literature as a test problem for unconstrained optimization techniques, due to its ill-conditioning. The function has a steep ridge over the set $\{(x_1, x_2) : x_2 - x_1^2 = 0\}$. The difficulty with optimizing this function is an example of the difficulty that general penalty function algorithms meet. Actually Rosenbrock's function can be interpreted as the following constrained problem:

$$\text{Minimize } (1-x_1)^2$$

Subject to

$$x_2-x_1^2=0.$$

If $I(x)=h^2(x)$, and a penalty parameter of 100 is used, the corresponding penalty function becomes Rosenbrock's function. Again, if we have a point (x_1, x_2) such that $x_2=x_1^2$, it is very hard to make any movement which will improve $f(x)+\lambda I(x)=(1-x_1)^2+100(x_2-x_1^2)^2$, since any improvement in the term $(1-x_1)^2$ will be offset by $100(x_2-x_1^2)^2$.

In spite of the above difficulties, efficient optimization of penalty functions can be accomplished. The following remarks are helpful in designing suitable unconstrained penalty optimization techniques.

(1) Starting with a small λ, and successively increasing it, has been reported in the literature as an efficient scheme (see, for example, [2]). Optimizing the penalty function with an initial small λ, usually leads to a "super optimal" point, i.e., an *infeasible* point with an objective smaller than the optimal objective. This is usually the case since a small initial λ automatically gives a more "relative weight" to the objective function. Therefore, rather than increasing λ via small steps, it is suggested that λ is increased sharply from its initial value, in order to get back to the feasible region as fast as possible.

(2) It does not pay off to do a complete optimization process with the chosen sequence of penalty parameters. For example, with the initial λ, a rough optimization is all that is needed. As λ increases, a tighter optimization is called for. This is implemented via the following simple *imperfect line search*. Given a function θ of one variable μ, an initial μ^* with $z^*=\theta(\mu^*)$, a step size $\Delta>0$, and an initial acceleration factor $s=1$, evaluate $\theta(\mu^*+s\Delta)$. If $\theta(\mu^*+s\Delta)<z^*$, then z^* is replaced by $\theta(\mu^*+s\Delta)$, μ^* is replaced by $\mu^*+s\Delta$, s is replaced by αs ($\alpha>1$, say $\alpha=2$), and the process is repeated, until the first failure is encountered, i.e., when $\theta(\mu^*+s\Delta)\geq z^*$. In this case, the *imperfect minimum* of θ is μ^*. Of course if the first trial results in a failure, then s is replaced by -1 and the process repeated.

Note that this "simple-minded" imperfect line search can essentially accomplish many objectives. First of all, during the first stages of optimization, when λ is small, Δ may be chosen fairly large, say Δ anywhere from 0.1 to 1. As the optimization process progresses and λ increases, Δ is decreased to allow moving along and close to the nonlinear ridge $\{x: I(x)=0\}$. Also note that the parameter s enables the search to progress quickly, if at all possible, even with very small values of Δ.

(3) Since the unconstrained penalty problem, with a large penalty parameter, is poorly behaved with elongated contours, one may be tempted to use a more "sophisticated" unconstrained minimization technique, in the hope of getting more accurate results. The opposite point of view is recommended. To illustrate, the possibility of using a conjugate direction based algorithm is discussed. If one decides to use conjugate directions to optimize a penalty function, one expects good results as long as the directions are truly conjugate (with respect to the Hessian at optimality), or closely so. However, in order to generate conjugate directions, usually information related to finding the true minimum of the penalty function along a certain direction is required. Due to ill-conditioning, this is almost impossible, and in view of compounding errors, the resulting directions are, in fact, far from being conjugate.

This results in more time spent per iteration, without significant pay-off. Of course using simple search techniques, in itself, will not help, unless the ability to move very small distances, and meanwhile being able to move an appreciable distance with the small stepsize via acceleration, is incorporated.

(4) One has the option of gradient versus nongradient search techniques. Since gradient and Hessian information require a significant amount of preparation, are usually subject to errors, and finally are very difficult to evaluate for certain classes of nonlinear programs, a nongradient search technique is recommended.

3. A CYCLIC COORDINATE ALGORITHM

In this section, and for the sake of completeness, a cyclic coordinate (one at a time) search procedure, for optimizing the penalty function is described. The procedure is the classical cyclic coordinate method, with minor modifications, which incorporates the remarks of the previous section.

The following notation and terminology are used.

d_i = ith coordinate axis, d_i is a vector of zeros, except one at the ith position.
y_{i+1} = n-dimensional vector obtained from y_i by optimizing along d_i.
Δ = initial stepsize.
$\Delta_{\min}$ = minimum stepsize used.
β = reduction rate of the stepsize.
λ = initial penalty parameter.
$\lambda_{\max}$ = maximum penalty parameter used.
γ = penalty acceleration factor.
ϵ = minimum displacement allowed per iteration.
t = iteration counter.
x_{t+1} = best available n-dimensional vector at the end of iteration t.
l = switching factor beyond which the penalty parameter and/or the stepsize are changed.
$|x|$ = denotes the sum of the absolute values of the x components.
$p(x, \lambda) = f(x) + \lambda I(x)$.

Initialization Step

Choose an arbitrary x_1 and choose the above parameters. Let $y_1 = x_1$, $i = t = 1$, and go to step 1.

STEP 1: Find the imperfect minimum of $p(y_i + \mu d_i, \lambda)$ with stepsize Δ. Let the minimum be μ^* and $y_{i+1} = y_i + \mu^* d_i$. If $i \leq n$, replace i by $i+1$ and repeat step 1. Otherwise go to step 2.

STEP 2: Let $x_{t+1} = y_{n+1}$. If $|x_{t+1} - x_t| \geq \epsilon$, let $i = 1$, $y_1 = x_{t+1}$, and repeat step 1. Now suppose that $|x_{t+1} - x_t| < \epsilon$. If $\lambda \geq \lambda_{\max}$ and $\Delta < \Delta_{\min}$, then stop, with optimal x_{t+1}. Otherwise check whether $\lambda\Delta$ is smaller than or greater or equal to l. In the former case, λ is replaced by $\gamma\lambda$ and in the latter case, λ is replaced by 2λ, and Δ is replaced by Max $(\beta\Delta, \Delta_{\min})$. Let $i = 1$, $y_1 = x_{t+1}$, and repeat step 1.

The above algorithm is self explanatory, perhaps with the exception of the switching factor l. If $\lambda\Delta < l$, then Δ is kept constant and the penalty parameter increased. On the other hand, if $\lambda\Delta \geq l$, then Δ is reduced, whereas λ is slightly increased. This device helps in executing incomplete minimization with small initial values of λ, and large values of Δ, while executing more precise minimization with small values of Δ, during later stages of the optimization.

Selection of the Parameters

In order to apply the above algorithm, several parameters must be chosen. Even though the optimal point and computational effort depend on the values of these parameters, our computational

experience with the procedure indicates that the results are not very sensitive to parameter selection. The following table may be helpful as a general guideline for the selection:

TABLE 1. *Parameter Values*

Parameter	Δ	Δ_{min}	β	λ	λ_{max}	γ	ϵ	l
Recommended	0.1 to 1.0	0.00001 to 0.001	0.1 to 0.4	0.1 to 1.0	100.0 to 1,000,000	5.0 to 100.0	0.0001 to 0.01	0.1 to 2.0
Range								

4. COMPUTATIONAL RESULTS

In this section, computational experience with the code is reported. Nine standard test problems have been attempted on a Univac 1108. The cyclic coordinate penalty method was able to solve all the test problems with optimal solutions equal to the optimal solutions reported in the literature. The problems attempted, include a variety of cases with nonlinear inequality, nonlinear equality, and mixed constraints, which shows that the method is "robust." Since no derivatives are used, the method can be used with minimal preparation time.

Since the procedure is based on peanlty functions, the optimal points obtained are feasible only within a certain amount of tolerance. Hence, a constraint of the form $h(x)=0$ (or $g(x) \leq 0$) may be evaluated at optimality as 0.02, in which case there is a violation of 0.02. The tolerance never exceeded 0.05 in any of the problems, and in most cases is much less than this value. The nine test problems were used by Stocker [7] (also see Himmelblau [3]) to test the relative efficiency of different nonlinear programming codes. Table 2 below shows a comparison between seven nonlinear programming codes and our proposed algorithm. The entries in the table, except for the entries of the proposed algorithm, are extracted from Table 9.3–6 of Himmelblau [3]. The following abbreviations are used in Table 2.

TABLE 2. *Comparison Between the Proposed Algorithm and Some Other NLP Codes*

Computational Time of Different Codes (sec)	Problem number[a]	*P*2	*P*1	*P*5	*P*8	*P*10	*P*7	*P*16	*P*11	*P*13
	Proposed algorithm[b]	0.07	0.12	0.16	0.31	1.4	0.90	0.58	1.25	0.63
	GRG	0.99	1.06	1.21	1.59	1.54	x	3.83	3.57	x
	FT	0.86	0.43	0.84	0.50	7.58	59.6	4.23	2.67	x
	NLP	0.18	0.07	0.33	14.04	1.62	x	1.60	2.31	x
	SUMT	0.11	0.51	x	1.80	2.79	x	12.66	1.05	x
	POP II	x	x	x	3.93	x	x	1.72	x	x
	Rosenbrock	0.21	4.52	x	0.70	x	x	0.52	1.71	x
	GGS	4.52	5.18	x	x	x	x	x	2.30	x

[a] The problem numbers correspond to the problem numbers of appendix A of Himmelblau [3].

[b] The proposed algorithm is executed on a Univac 1108, whereas other results correspond to CDC 6600.

x—means that the code cannot be applied to the problem or the code was not able to solve the problem.

For more details about these codes, the reader may refer to [3] and references therein.

GRG: Generalized Reduced Gradient

FT: Flexible Tolerance

NLP: Nonlinear Programming

SUMT: Sequential Unconstrained Minimization Technique

POP: Process Optimization Program

Rosenbrock: Rosenbrock's procedure, accommodating inequality constraints by Logarithmic penalty

GGS: Generalized Gradient Search

REFERENCES

[1] Bazaraa, M. S., "Geometry and Resolution of Duality Gaps," Nav. Res. Log. Quart. *20*, 357–365, 1973.

[2] Fiacco, A. V. and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (John Wiley and Sons, Inc., 1968).

[3] Himmelblau, D. M., *Applied Nonlinear Programming* (McGraw-Hill Book Company, 1972).

[4] Lasdon, L. S., "An Efficient Algorithm for Minimizing Barrier and Penalty Functions," Mathematical Programming (1972), vol. 2.

[5] Luenberger, D. G., *Introduction to Linear and Nonlinear Programming* (Addison-Wesley Publishing Company, 1973).

[6] Rosenbrock, H. H., "An Automatic Method for Finding the Greatest or the Least Value of a Function," Comput. J. (1960), vol. 3.

[7] Stocker, D. C., "A Comparative Study of Nonlinear Programming Codes," M. S. Thesis, The University of Texas, Austin, Texas (1969).

[8] Zangwill, W. I., *Nonlinear Programming: A Unified Approach* (Prentice-Hall, 1969).

ON FRACTIONAL PROGRAMMING AND EQUIVALENCE

B. D. Craven

University of Melbourne
Melbourne, Australia

and

B. Mond

La Trobe University
Melbourne, Australia

ABSTRACT

Under fairly general conditions, a nonlinear fractional program, where the function to be maximized has the form $f(x)/g(x)$, is shown to be equivalent to a nonlinear program not involving fractions. The latter program is not generally a convex program, but there is often a convex program equivalent to it, to which the known algorithms for convex programming may be applied. An application to duality of a fractional program is discussed.

1. EQUIVALENT FRACTIONAL PROGRAMS

The two mathematical programs

(I) Maximize $f(x)$ subject to $x \in \mathbf{S}$,

(II) Maximize $F(\xi)$ subject to $\xi \in \mathbf{T}$,

are called q-equivalent (see [3]) if there is a one-one map q of the constraint set $\mathbf{S}$ of (I) onto the constraint set T of (II), such that $f(x) = F(q(x))$. Then q-equivalence is reflexive and transitive; and it follows readily from q-equivalence, ([3], Lemma 1) that if (I) attains a (global) maximum at x^*, then (II) attains a (global) maximum at $q(x^*)$; and if (II) attains a (global) maximum at ξ^*, then (I) attains a (global) maximum at $q^{-1}(\xi^*)$. Similar definitions apply to minimizing programs.

Consider the fractional programming problem

(P) Maximize $f(x)/g(x)$ subject to $h(x) \in \mathbf{S}$ and $g(x) > 0$,

where $x \in \mathbf{R}^n$, and $f: \mathbf{R}^n \to \mathbf{R}$, $g: \mathbf{R}^n \to \mathbf{R}$, and $h: \mathbf{R}^n \to \mathbf{R}^m$ are given maps, and $\mathbf{S}$ is a convex cone in $\mathbf{R}^m$. Let $\phi_0: \mathbf{R} \to \mathbf{R}$ be a strictly increasing function, with $\phi_0(t) > 0$ for $t > 0$. Let T be a convex cone in $\mathbf{R}^s$ and, for each $t \in \mathbf{R}_+ = [0, \infty)$, let $\psi(t): \mathbf{R}^m \to \mathbf{R}^s$ be a map, such that $z \in S \iff \psi(t)[z] \in T$. (A particular case is $S = T = \mathbf{R}^m_+$, and $\psi(t)$ a diagonal matrix with positive diagonal elements for each t.) For $t \in \mathbf{R}_+$ and $y \in \mathbf{R}^n$, define

$$F(y, t) = f(y/t)\phi_0(t); \qquad G(y, t) = g(y/t)\phi_0(t); \qquad H(y, t) = \psi(t)[h(y/t)].$$

Define also

$$F(y, 0) = \lim_{t \to 0+} F(y, t); \qquad G(y, 0) = \lim_{t \to 0+} G(y, t); \qquad H(y, t) = \lim_{t \to 0+} H(y, t),$$

whenever these limits exist. Assume that $G(0, 0) = 0$ whenever it exists.

Consider the associated nonlinear programming problem

(P*) Maximize $k^{-1}F(y, t)$ subject to $G(y, t) = k$ and $H(y, t) \in T$;

where k is a positive constant.

THEOREM 1: If, for each $y \in \mathbf{R}^n$, $(y, 0)$ is not a feasible point for (P*), then the problems (P) and (P*) are q-equivalent, with q defined by

$$q(x) = (y, t), \qquad \text{where} \qquad t = \phi_0^{-1}\ (k/g(x)) \qquad \text{and} \qquad y = tx.$$

REMARK: Consequently, if either of (P) and (P*) has a maximum, then so has the other, and the maxima are equal.

Theorem 1 generalizes the results of [5] and [6], in which $S = R^m_+$; in [6], the constraints also were restricted to be linear.

PROOF: If $x \in \mathbf{R}^n$ is feasible for (P), then $(y, t) = q(x) \in \mathbf{R}^n \times \mathbf{R}_+$ is defined uniquely, with $t > 0$; since the case $t = 0$ is excluded, the inverse map is given by $x = q^{-1}(y, t) = y/t$, and so q is one-one; $h(x) \in S \Rightarrow H(y, t) \in T$; and $g(x) > 0$ and $t = \phi_0^{-1}\ (k/g(x)) > 0 \Rightarrow G(y, t) = k$. If $(y, t) \in \mathbf{R}^n \times \mathbf{R}_+$ is feasible for (P*), then $q^{-1}(y, t) \in \mathbf{R}^n$, since $t > 0$ by hypothesis; $g(x) = G(tx, t)/\phi_0(t) > 0$; and $H(y, t) = \psi(t)[h(x)]$, with the hypothesis on $\psi(t)$, implies that $h(x) \in S$. So q maps the constraint set of (P) one-one onto the constraint set of (P*).

Since also

$$k^{-1}F(y, t) = \frac{F(y, t)}{G(y, t)} = \frac{f(y/t)\phi_0(t)}{g(y/t)\phi_0(t)} = \frac{f(x)}{g(x)},$$

the map q fulfils the requirements for q-equivalence.

REMARKS: In the particular case of linear constraints, when h is an affine function, then $t > 0$ follows ([1]) from the hypothesis that (P) has a bounded constraint set.

In general, if $(y, 0)$ is a feasible point for (P*), then either (i) $(y, 0)$ is a limit of feasible points (y, t) with $t > 0$, or (ii) this is not the case. Suppose that (P) has a bounded constraint set; thus

$$g(x) > 0 \qquad \text{and} \qquad h(x) \in S \Rightarrow x^T x \leq M < \infty.$$

If $t > 0$ then, since $x = y/t$,

$$G(y, t) = k \qquad \text{and} \qquad H(y, t) \in T \Rightarrow g(x) > 0 \qquad \text{and} \qquad h(x) \in S \Rightarrow x^T x = t^{-2} y^T y \leq M.$$

So if $(y, 0)$ is feasible of kind (i), then, letting $t \downarrow 0, y^T y \leq Mt^2 \downarrow 0$, so $y = 0$. This gives a contradiction if k is chosen so that $k \neq G(0, 0)$.

2. EQUIVALENT CONVEX PROGRAM

Since the constraint set of (P*) is not generally a convex set, even if the functions G and $-H$ are convex functions of (y, t), (P*) may be no easier to solve than the equivalent fractional program (P). If G is a convex function of the vector variable (y, t), then the set $\{(y, t) : G(y, t) \leq k, y \geq 0, t > 0\}$ is convex, but the corresponding set with $G(y, t) = k$ is not generally convex.

Theorem 2 shows that a programming problem whose constraint set is not convex can sometimes be replaced by an "equivalent" problem with a convex constraint set. Here, "equivalent" does not mean q-equivalent, since the constraint set is not mapped one-one, but merely that the "equivalent" problem has the same maximum at the same point as the given problem.

THEOREM 2: If p is a convex function which is locally-bounded (i.e., bounded in a neighbourhood of each point), then the problems

(P2) Maximize $p(z)$ subject to $z \in \mathcal{W}$

(where $\mathcal{W} \subset \mathbf{R}^r$), and

(P2′) Maximize $p(z)$ subject to $z \in \overline{co}\,\mathcal{W}$

(where $\overline{co}$ denotes closed convex hull) attain the same maximum at the same point.

PROOF: Let (P2) attain its maximum at the point $z = z^*$; then $p(z) \leq p(z^*)$ for each $z \in \mathcal{W}$. Since $\mathcal{W} \subset \overline{co}\,\mathcal{W}$, it is sufficient to prove that $p(z) \leq p(z^*)$ for each $z \in \overline{co}\,\mathcal{W}$. Each point $z' \in co\,\mathcal{W}$ (the convex hull of $\mathcal{W}$) is expressible as a finite sum $z' = \Sigma \alpha_i z_i$, where each $\alpha_i \geq 0$, $\Sigma \alpha_i = 1$, and each $z_i \in \mathcal{W}$. Since p is convex,

$$p(z') \leq \Sigma \alpha_i p(z_i) \leq (\Sigma \alpha_i) p(z^*) = p(z^*).$$

If $z \in \overline{co}\,\mathcal{W}$, then z is a limit of points $z' \in co\,\mathcal{W}$; since p is convex and locally-bounded, it is continuous ([2], page 326), so $p(z) \leq p(z^*)$.

REMARKS: The proof shows that z^* is maximal for (P2′), as well as for (P2). The possibility is not excluded that (P2′) may have an additional maximum $z^{**} \in (\overline{co}\,\mathcal{W}) - \mathcal{W}$, which is not a maximum of (P2), for which $p(z^{**}) = p(z^*)$.

Theorem 2 can be put to use if $\overline{co}\,\mathcal{W}$ can be specified by a finite number of inequality constraints. This occurs for (P*), under suitable additional hypotheses.

DEFINITION: A function $G : \mathbf{R}^r_+ \to \mathbf{R}$ will be called *compatible* if, for a fixed $k > 0$, G is continuous, $G(0) < k$, $Y = \{z \in \mathbf{R}^r_+ : G(z) = k\}$ is bounded, Y intersects each coordinate axis in $\mathbf{R}^r_+$, and the sets

$$E = \{z \in \mathbf{R}^r_+ : G(\mathbf{z}) \leq k\} \quad \text{and} \quad Q = \{z \in \mathbf{R}^r_+ : G(z) < k\}$$

are convex.

REMARKS: If G is a convex function, then E and Q are convex sets; but the converse does not hold. In general, Y is not a convex set.

Since G is continuous, $G(0) < k$, and Y intersects the jth coordinate axis Z_j,

$$0 < b_j = \inf (Y \cap Z_j) < \infty.$$

In Theorem 3, $\mathbf{R}^r = \mathbf{R}^n \times \mathbf{R}$, with $z = (y, t)$; the corresponding numbers b_j will be written b_1, b_2, . . ., b_n, b_0; let d be the column vector whose components are b_1^{-1}, b_2^{-1}, . . ., b_n^{-1}, b_0^{-1}.

LEMMA: If G is compatible, then the closed convex hull

$$\overline{co}\, Y = E \cap H_+,$$

where H_+ is the half-space

$$H_+ = \{z = (z_1, \ldots, z_r) \in \mathbf{R}^r : \sum_{j=1}^{r} (z_j/b_j) \geq k\}.$$

PROOF: If $z \in \mathbf{R}_+^r - H_+$ then, for some $\epsilon > 0$, z is a point of the simplex M whose vertices are 0 and the points $(0, 0, \ldots, 0, b_j - \epsilon, 0, \ldots, 0)$, for $j = 1, 2, \ldots, r$. But each of these vertices lies in Q; since Q and M are convex, $z \in M \subset Q$, so $z \notin Y$; hence $Y \subset H_+$. So $Y \subset J$, where $J = E \cap H_+$; since H_+ and E are each closed convex sets, $\overline{co}\, Y \subset J$.

Conversely, let $z \in J$. Since Y is closed and bounded, and G is continuous, the ray from 0 through z contains a point $u \in Y$ of maximal Euclidean norm. If $G(z) = k$, then $z \in Y \subset co\, Y$. If $G(z) < k$, and u lies in the segment between 0 and z, then, since $G(0) < k$ and E is convex, $G(u) < k$, contradicting $u \in Y$. Hence z lies in the segment L joining 0 to u. The point where L intersects the boundary of H_+ is a point of $co\, Y$, and z is then a convex combination of this point and u, so $z \in co\, Y$. Thus $J \subset co\, Y$.

DEFINITION: The function $H : \mathbf{R}^r \to \mathbf{R}^s$ is concave with respect to the cone $T \subset \mathbf{R}^s$ if, for each $\lambda \in (0, 1)$ and each $z_1, z_2 \in \mathbf{R}^r$,

$$H(\lambda z_1 + (1-\lambda) z_2) - \lambda H(z_1) - (1-\lambda) H(z_2) \in T.$$

THEOREM 3: Let F, G, and H be the functions defined in the problem (P*), and let F be convex, G compatible (with positive constant k), and H concave with respect to the convex cone T. If, for each $y \in \mathbf{R}_+^n$, the point $(y, 0)$ is not feasible for the problem

(P**) Maximize $k^{-1} F(y, t)$

subject to

$$G(y, t) = k, \qquad H(y, t) \in T, \qquad y \in \mathbf{R}_+^n, \qquad t \in \mathbf{R}_+;$$

then (P**) attains the same maximum at the same point (y^*, t^*) as the convex programming problem

(CP) Maximize $k^{-1} F(y, t)$

subject to

$$G(y, t) \leqslant k, \qquad H(y, t) \in T, \qquad y \in \mathbf{R}_+^n, \qquad t \in \mathbf{R}_+,$$

and

$$\sum_{j=1}^{n} \frac{y_j}{b_j} + \frac{t}{b_0} \geqslant k.$$

PROOF: Apply Theorem 2 and the Lemma.

3. APPLICATIONS

EXAMPLE 1: Let γ, δ, $\epsilon_i \in \mathbf{R}$; a, d, $c_i \in \mathbf{R}^n$; C, $D_i \in \mathbf{R}^{n \times n}$ $(i = 1, 2, \ldots, m)$; and let the matrices C and D_i be symmetric positive semidefinite. Define the polyhedral cone $S \subset \mathbf{R}^m$ as $S = Q^{-1}\mathbf{R}_+^s = \{w \in \mathbf{R}^m \cdot Qw \in \mathbf{R}_+^s\}$, where $Q : \mathbf{R}^m \to \mathbf{R}^s$ is a linear mapping, which maps S one-one onto $\mathbf{R}_+^s$. Consider the problem (P) with

$$f(x) = a^T x + \gamma; \qquad g(x) = d^T x + (x^T C s)^{1/2} + \delta;$$

$h = (h_1, h_2, \ldots, h_m)$, where $h_i(x) = c_i^T x - (x^T D_i x)^{1/2} - \epsilon_i$; and the additional constraint $x \in \mathbf{R}_+^n$. Assume that the constraint set of (P) is bounded.

If $\phi_0(t) = t$ and $\psi(t) = tQ$, then $F(y, t) = a^T y + \gamma t$ is linear;

$$G(y, t) = d^T y + (y^T C y)^{1/2} + t\delta$$

is convex, since a linear function is convex, and $(y^T C y)^{1/2}$ defines a convex norm; similarly, for $i = 1, 2, \ldots, m$, $H_i(y, t)$ is concave, so that $H = (H_1, \ldots, H_m)$ is concave with respect to

$$T = \psi(t)[S] = tQS = t\mathbf{R}_+^m = \mathbf{R}_+^m.$$

If $(y, 0)$ is not feasible for (P**), then Theorem 3 may be applied to this example, to generate an equivalent convex programming problem.

Similar conclusions apply if, instead, $h_i(x) = -x^T D_i x - \beta_i$, where D_i is positive semidefinite and $\beta_i > 0$; here, if $\psi(t) = t^2 Q$, then also H is concave with respect to $T = \mathbf{R}_+^m$. In this case, if the constraint set for (P) is nonempty and bounded, then $(y, 0)$ is not feasible for (P**) (see [5]).

EXAMPLE 2: Denote by $(\bar{\mathrm{P}})$ the problem (P), with additional constraint that $x \in \mathbf{R}_+^n$, under the hypotheses that, in the corresponding problem $(\bar{\mathrm{P}}^*)$, F is linear (thus $F(y, t) = p^T z$ where $z = (y, t)$ and $P \in \mathbf{R}^{n+1}$), G is compatible, H is concave with respect to $T = \mathbf{R}_+^m$, and G and H are differentiable. Assume also that $(y, 0)$ is not feasible for the corresponding problem (P**); then the hypotheses of Theorem 3 are fulfilled. If the Kuhn-Tucker constraint qualification applies to (CP) then, since F is concave, the usual duality theorem for convex programming ([4], page 115) shows that the dual program to (CP) is

(D) Minimize $-u^T[G(z) - k - G'(z)z] + v^T[H(z) - H'(z)z] - ks$

subject to $z \in \mathbf{R}_+^r, \quad u \in \mathbf{R}_+, \quad v \in \mathbf{R}_+^m, \quad s \in \mathbf{R}_+,$

$$u^T G'(z) - v^T H'(z) - k^{-1} p^T - s d^T \in \mathbf{R}_+^r;$$

where the symbols $F'(z)$, etc., denote Fréchet derivatives.

As in [3], the program

(1) Minimize $\phi(w)$ subject to $w \in U$,

is a *strong dual* of the program

(2) Maximize $\psi(x)$ subject to $x \in Y$,

if (a) (1) has the optimal solution $w^* \Longleftrightarrow$ (2) has an optimal solution x^*, and $\psi(x^*) = \phi(w^*)$, and (b) if $x \in Y$ and $w \in U$, then $\psi(x) \leqslant \phi(w)$.

It follows from Theorems 1, 2, and 3 and the usual duality theorem that (D) is a strong dual of the nonlinear fractional programming problem $(\overline{\mathrm{P}})$.

REFERENCES

[1] Charnes, A. and W. W. Cooper, Programming With Linear Fractional Functionals, Nav. Res. Log. Quart. *9*, 181–186 (1962).

[2] Courant, R., *Differential and Integral Calculus* (Blackie, London, 1936), vol. II.

[3] Craven, B. D. and B. Mond, "The Dual Of A Fractional Linear Program," Journal of Mathematical Analysis and Applications *42*, 507–512 (1973).

[4] Mangasarian, O., *Nonlinear Programming* (McGraw-Hill, New York, 1969).

[5] Mond, B. and B. D. Craven, "Nonlinear Fractional Programming," Bull. Austral. Math. Soc. *12*, 391–397 (1975).

[6] Mond, B. and B. D. Craven, A Note On Mathematical Programming With Fractional Objective Functions, Nav. Res. Log. Quart. *20*, 577–581 (1973).

A NOTE ON CONTROL THEORY AND ATTAINING EQUILIBRIUM

Charles Tiplitz
Chalon Consultants
Potomac, Md.

ABSTRACT

Most researchers have not recognized that the main problem of an economic system is the excessive time it requires to reach equilibrium (and consequent inadequate control). An extension of control theory is applied to the Competitive Equilibrium Model. The calculation shows, in contrast to previous results, full employment, if attained, may be maintained for nonzero periods. However, ill condition and other factors cause a very slow approach to optimum. Consequently, constant social and other changes keep a complex economy far from optimum.

The competitive equilibrium model has been much maligned[1] and misunderstood. Yet the main problem, universally overlooked, is the *time* rate of convergence or response rate.[2] The likely ill-condition of large matrices,[3] which characterize complex economies, shows that the economy of a developed nation has a long time constant. It is this long time constant (and the consequent excessively loose control response) that really explains economic stabilization problems and reconciles the logic of the law of supply and demand with its unsatisfactory resource allocation results.

Furubotn discusses the ethical differences between economic options and apparently assumes the time to go from one to the other is a problem of no consequence.[4] Lele, Jacobson and McCabe[5] applied control theory to several econometric models of economic growth to show that optimal trajectories do not include nonzero boundary arcs and, in particular, that full employment could not exist continuously. They, too, overlooked the time probably required to reach equilibrium. The theory

---

[1]Nicholas Kaldor, ("The Irrelevance of Equilibrium Economics," *Economic Journal*, September–December 1972) writes off. However, the equilibrium model is based on the logic of the law of supply and demand, identifying an optimum which is a valid target even in the absence of competition. All the empirical evidence against the required mathematical conditions show, at best, minor aberrations from the fundamental relations. Linearity is often considered a serious defect although linearity is not an essential characteristic. Linear relations are simpler to manipulate which is the reason for their occurrence. In many cases nonlinear effects are second order and do not alter significantly conclusions based on linear calculations.

[2]All the literature and research on equilibrium deal with whether there is convergence (ignoring time) to a unique solution and the absolute and precise condition under which convergence occurs.

[3]A. Ralston, *A First Course in Numerical Analysis*, (New York: McGraw-Hill, 1965), p. 234.

[4]Eirik G. Furubotn, "Economic Organization and Welfare Distribution," *Swedish Journal of Economics*, Vol. 73, No. 4, December 1971, p. 409–416.

[5]Milind M. Lele, et al., "Qualitative Application of a Result in Control Theory to Problems of Economic Growth," *International Economic Review*, Vol. 12, No. 2, June 1971, p. 209–226. Masanao Aoki (*Review of Economic Studies*, Vol. 40, January 1973, p. 131–137) applied control theory to an econometric model to show what formal sufficient conditions are, but did not consider the underlying economic system.

applied, was the refinement by Jacobson and Lele of optimality conditions for control problems with inequality boundaries.[6] Lele, et al., explain that the policy implications of their models suggested more direct control.[7]

However, such direct control is embodied in the Competitive Equilibrium Model. Therefore, had control theory been applied to that model, results would be different as follows. Referring only to the relevant equations as presented by Karlin[8] we have the objective function:

(1) $$\text{Maximize } (P^*, X)$$

where P^* is the equilibrium price vector whose components represent prices of a set of products, X is a vector representing consumption, Y^* is a vector representing production at equilibrium, and $X \leqslant Y^*$ the control function assumed for a competitive market:

(2) $$\frac{dP}{dt} = X - Y,$$

where Y is production, and the state constraint:

(3) $$AY \leqslant B$$

where A is the matrix of technological coefficients, and B are the given resource constraints.

Now, Equation (3), the state constraint, contains Y and hence from Equation (2), the control variable, $\dot{Y}$, is explicitly in the first derivative of Equation (3). Therefore, the state constraint is of first order. This implies that the optimum trajectory for a market economy may include nonzero boundary arcs, that is, arcs corresponding to the equality in Equation (3). In other words, full employment may be maintained continuously if labor is one of the resources. The exact trajectory can be determined only if the data were available, but an indication may be obtained by use of time constants, assuming fixed constraints and tastes.

From such analysis it will be seen that a more critical problem is the initial approach to the boundary. Equation (2) is first expanded in a Taylor Series about the equilibrium point to obtain:

(2a) $$\frac{dP}{dt} = (X^* - Y^*) + \frac{\partial (X - Y)}{\partial P}(P - P^*) + \ldots,$$

where X^* is the equilibrium consumption and P is price.

From Walras' Law ($(P^*, X^* - Y^*) = 0$) it follows that $X^* = Y^*$. Also $\frac{dP^*}{dt} = 0$. Therefore, we may write:

---

[6] D. H. Jacobson, M. M. Lele and J. L. Speyer, "New Necessary Conditions of Optimality for Control Problems with State-Variable Inequality Constraints," *Journal of Mathematical Analysis and Applications*, Vol. 35, August 1971, p. 255–284.

[7] Op. Cit., p. 219.

[8] S. Karlin, *Mathematical Methods and Theory in Games, Programming, and Economics* (London: Pergamon Press, 1959) p. 293–301.

4) $$d\hat{P} \cong \frac{\partial (X-Y)}{\partial P} \hat{P},$$

where $\hat{P} = P - P^*$. Then the solution is:

5) $$\hat{P}_i = \hat{P}_{i0} e^{\lambda_i t}$$

where λ_i is the ith eigenvalue of the matrix $\frac{\partial (X-Y)}{\partial P}$ and $\hat{P}_{i0}$ is the corresponding eigenvector. Now, since large matrix (chracteristic of complex societies) are most often ill-conditioned, it follows that at least one eigenvalue is very small, i.e., the time constant is excessively long (Sato speaks of 15 years based on empirical studies).[9] This implies that the boundary may be approached so slowly, that social, technological and political changes will keep the optimum forever far from reach.

A simple two commodity example will illustrate the problem. The matrix $\frac{\partial (X-Y)}{\partial P}$ is the matrix of cross elasticities:

$$\begin{pmatrix} \frac{\partial (x_1 - y_1)}{\partial P_1} & \frac{\partial (x_2 - y_2)}{\partial P_1} \\ \frac{\partial (x_1 - y_1)}{\partial P_2} & \frac{\partial (x_2 - y_2)}{\partial P_2} \end{pmatrix}.$$

Normally, the diagonals dominate, and are negative reflecting the adverse impact of a rise in price on consumption. Assume the ill-conditioned matrix,

$$\begin{pmatrix} -1 & 0.8 \\ 0.8 & -1 \end{pmatrix},$$

and time, t, in Equation (2) is measured in years. Then the eigenvalues are

$$\lambda = -1.8, -0.2$$

and the respective eigenvalues are

$$P_{10} = (p, -p)$$

$$P_{20} = (p, p).$$

---

[9] Cf. K. Sato, "On the Adjustment Time in Neo-Classical Growth Models," *Review of Economic Studies*, 33 (1966), pp. 263–268, where the time to reach a stable growth rate is calculated at 15 years. If Equation (2) is violated, such as by institutionalization (prices can rise when airline seats are empty, etc.), then some eigenvalue may even have positive real parts (i.e., the system may even be unstable).

Since there are two eigenvectors Equation (5) becomes in this illustration

$$\hat{P} = a\hat{P}_{10}e^{-1.8t} + b\hat{P}_{20}e^{-0.2t}$$

where a and b are constants dependent on initial conditions. Assume that an instantaneous drop in demand from equilibrium occurs for one component

$$x_1 - y_1 = -0.05,$$

$$x_2 - y_2 = 0,$$

and calculate the response. Differentiating the expression for $\hat{P}$ and setting its components equal to the above initial condition as follows:

$$\begin{pmatrix} -.05 \\ 0 \end{pmatrix} = -1.8a \begin{pmatrix} p \\ -p \end{pmatrix} - 0.2b \begin{pmatrix} p \\ p \end{pmatrix},$$

yields, for this initial condition plus the initial condition $\hat{P} = 0$:

$$\hat{P}(t) = \begin{pmatrix} .0138 \\ -.0138 \end{pmatrix} e^{-1.8t} + \begin{pmatrix} .125 \\ .125 \end{pmatrix} e^{-0.2t} - \begin{pmatrix} .1388 \\ .1112 \end{pmatrix}, \text{ and } \dot{P}(t) = \begin{pmatrix} -.025 \\ .025 \end{pmatrix} e^{-1.8t} + \begin{pmatrix} -.025 \\ -.025 \end{pmatrix} e^{-0.2t}.$$

At the end of 1 year:

$$\dot{P}(t) = \begin{pmatrix} -.0041 \\ .0041 \end{pmatrix} + \begin{pmatrix} -.0205 \\ -.0205 \end{pmatrix} = \begin{pmatrix} -.0246 \\ -.0164 \end{pmatrix}.$$

We retain the sum in the calculation to show that the ill-condition slowed the adjustment. The first vector in the sum decreased to one-sixth its initial value. But the second vector, corresponding to the second and low eigenvalue, dropped by only 20 percent, indicating only a small adjustment took place in the first year. Note that demand for the second commodity has dropped also because of the cross elasticity.

Lele, et al., discuss research possibilities, but they and others seem to have overlooked the opportunity and need regarding time rate of convergence.

☆U.S. GOVERNMENT PRINTING OFFICE: 1975-211-271/4

INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Arlington, Va. 22217. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 250 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.

NAVAL RESEARCH LOGISTICS QUARTERLY

JUNE 1975
VOL. 22, NO. 2
NAVSO P-1278

CONTENTS

ARTICLES

		Page
Dynamic Multiattribute Models for Mixed Manpower Systems	A. CHARNES, W. W. COOPER, R. J. NIEHAUS	205
Insurance Type Item Provisioning Guidelines	R. W. BURTON, S. C. JAQUETTE	221
Periodic Replacement with Minimal Repair at Failure and Adjustment Costs	C. TILQUIN, R. CLEROUX	243
Optimal Schedules for Special Structure Flowshops	J. N. D. GUPTA	255
Sensitivity of Queueing Models to the Assumption of Exponentiality	D. GROSS	271
Inference from Accelerated Life Tests Using Eyring Type Re-Parameterizations	N. D. SINGPURWALLA, V. C. CASTELLINO, D. Y. GOLDSCHEN	289
An Operator Theory of Parametric Programming for the Generalized Transportation Problem—III—Weight Operators	V. BALACHANDRAN, G. L. THOMPSON	297
An Operator Theory of Parametric Programming for the Generalized Transportation Problem—IV—Global Operators	V. BALACHANDRAN, G. L. THOMPSON	317
A Recursive Branch and Bound Algorithm for the Multidimensional Knapsack Problem	A. THESEN	341
Parametric Studies in Transportation-Type Problems	C. O. FONG, M. R. RAO	355
Adjacent Vertices on Transportation Polytopes	P. G. MCKEOWN, D. S. RUBIN	365
The Static Theory of Transfer Pricing	H. ENZER	375
Isoquants of Continuous Production Correspondences	G. BOL, O. MOESCHLIN	391
An Efficient Cyclic Coordinate Method for Optimizing Penalty Functions	M. S. BAZARAA	399
On Fractional Programming and Equivalence	B. D. CRAVEN, B. MOND	405
A Note on Control Theory and Reaching Equilibrium	C. I. TIPLITZ	411

OFFICE OF NAVAL RESEARCH
Arlington, Va. 22217